The Genetics of familial non-specific intellectual disability in the Kuwaiti population

Al-Mosawi, Fatemah Hashem

Awarding institution:
King's College London

The Genetics of familial non-specific intellectual disability in the Kuwaiti population

Fatemah Hashem Al-Mosawi

Submitted for the degree of Doctor of Philosophy in Genetics

King's College London, University of London

Sep 2013

Abstract

Detection of the genetic causes of non-specific intellectual disability (ID) and many neurodevelopmental disorders is a complex process, yet this information has a great impact on aiding clinical geneticists to inform families with affected members or carrier status for better decision making. Advances in technology allow the detection of smaller aberrations such as copy number variations (CNV), which have been linked to several disorders. This study aims to establish a diagnostic workflow in the Kuwait Medical Genetic Centre for heterogeneous disorders such as ID and a rare syndrome, Rubinstein-Taybi Syndrome (RTS). In 60% of RTS cases, the genetic cause is a mutation in one of two genes, with the cause unknown for the remaining cases. Different gene detection methods were evaluated in individuals from 16 consanguineous families with multiple affected sibs with non-specific ID and 10 RTS individuals. Laboratory methods of direct gene sequencing using the ABI sequencer, multiple probe-dependent ligation assay (MPLA) and the Affymetrix 2.7M cytogenetic arrays for analysis of CNVs and runs of homozygosity were used in these individuals. In addition, one family was studied further through homozygosity mapping and exome sequencing. Further confirmatory testing is required but the study produced several recommendations for the future direction of Clinical Genetic diagnosis in Kuwait as well as specific follow up options on the potential causative CNVs.

Acknowledgement

I would like to thank my mother Bushra AlKazemi, for always believing in me and supporting me to pursue my dreams, my family and friends for their continual support and encouragement, my supervisors for their advice and namely Cathryn Lewis who has been amazing at keeping me on track. Francesca Ducci for her support with the exome sequencing, Shaza Al-Sabban, Dafina Shabani, and Louise Arseneault for the student support they offered me.

Members of the Kuwait Medical Genetics Centre, staff and colleagues, namely Dr Sadika Al-Awadi who always encouraged me and opened opportunities for me to pursue research, Dr Makia Marafie, the amazing Clinical Genetics who was like a God mother to me.

My family at Spread the Passion and especially Dana AlEssa and colleagues at the National Youth Project who believed in me, namely Dr Yousef AlEbraheem, Jasem AlRabaian, Abdullah AlBader, Abdulaziz Sadeq, and Shamael AlSharikh.

All my friends who kept cheering me on namely, Dr Winston Wai Shing Lau and Abdulkareem AlSaqer for being there for me and keeping me focused.

Finally, this work is dedicated to my beloved husband Mohamad Mahdi and two sons my Pride and Joy, Jaffer and Hamza, as they had to bear with me during the course of my studies and I hope my achievement would inspire them to do the same and surpass me.

Statement of work

This project was funded through the bench fees of the scholarship from the Ministry of Health of Kuwait through the civil service commission. Some reagents were supplied directly from the Ministry of Health of Kuwait. Laboratory work was carried out in the Kuwait Medical Genetic Centre and the Core shared facilities at the Medical School in Kuwait University. The author carried out all laboratory work, with minimal training and following protocols listed in manufacturer's manuals, with the exception of some DNA extraction for some samples that were carried out by other laboratory technicians and two samples were sent off for exome sequencing at Guy's hospital.

The purpose of this work is to develop the Clinical Genetic service by improving the diagnostic capabilities of the Kuwait Medical Genetic Centre using two disorders as working examples.

Table of Content

List of Tables

List of figures

Chapter1-Introduction

Background

Health service is an essential commodity in modern life; this encapsulates a number of diagnostic and clinical services, which include clinical genetic services. A growing number of health clinics are developing their clinical genetic capabilities to offer better service, as further understanding of genetics is shaping the future of medicine. This has opened the diagnostic potentials for many disorders that were not very well understood in the past.

Likewise, Kuwait Medical Genetic Centre is developing such services to provide better quality and high assurance service to the Kuwaiti and expatriate population. This includes exploring options for the introduction of new diagnostic tools and protocols, and the investment in new laboratories to detect the underlying genetic cause of more disorders.

Moreover, the large family size of those affected by genetic disorders, consanguineous nature and relatedness of many of the marriages in Kuwait, and the introduction of premarital counselling requires better diagnostic tools to complement the clinical genetic services on offer.

Furthermore, the workflow and advances of gene detection has been investigated in this study using two disorders, non-specific familial intellectual disability (ID) and Rubinstein-Taybi syndrome (RTS) a rare disorder.

This chapter covers the background on intellectual disability, RTS, types of genetic variation, modes of inheritance, approaches to identifying genetic causes, laboratory techniques to variation detection, and the current clinical genetic services offered in Kuwait which lead to the purpose of this study.

1.1 Intellectual disability

Intellectual disability (ID), also known as mental retardation (MR) and sometimes referred to as cognitive impairment (Ropers 2008; Ellison, Rosenfeld et al. 2012; Kakar, Goebel et al. 2012; Mefford, Batshaw et al. 2012) is the most common neurodevelopmental disorder and is defined as a reduced ability to learn, marked by an Intelligence quotient (IQ) score of below 70, and an impairment in at least two adaptive behaviour independently, both of which must be presented before the age of 18 (Kaufman, Ayub et al. 2010; Ellison, Rosenfeld et al. 2012; Soltani Banavandi, Kahrizi et al. 2012).The term ID will be used throughout this document to refer to the disorder, as there is an international consensus on the use of this term rather than mental retardation, although for the purpose of literature search the previous term would yield more results.

ID is a lifelong disability and is thought to affect 1-3 % of the general population (Galasso, Lo-Castro et al. 2010). It can be further divided into categories based on IQ level (DSM-IV classification) that includes, borderline (IQ score 70-84), mild (IQ score 50-69), moderate (IQ score 35-49), severe (IQ score 20-34) and profound (IQ score below 20). The prevalence of severe ID is thought to have a strong genetic component where as mild to moderate ID can be the result of or the interaction of other

environmental component including maternal level of health education (Kaufman, Ayub et al. 2010). The American Psychiatric Association reported a 30% higher incidence of ID in males compared to females in mild cases (which could be due to X-linked causes). However in severe cases the incidence is thought to be higher amongst women. This is based on specific communities and an overall association between gender and the incidence of ID is difficult to establish. Incidence is thought to be higher in individuals with lower socioeconomic status and in developing countries, which could be attributed to environmental factors or consanguineous marriages (Kaufman, Ayub et al. 2010).

1.1.1 Causes of ID

ID is a very heterogeneous disorder with a mixed aetiology. In more than half of cases the cause remains unknown. Causes can include environmental factors that damage or interfere with normal brain development including pre- or postnatal insults such as exposure to toxins, teratogens, radiation, alcohol, lead, malnutrition, infections, trauma, and injury causing lack of oxygen to the brain (Kaufman, Ayub et al. 2010; Ellison, Rosenfeld et al. 2012; Mefford, Batshaw et al. 2012).

Genetic factors account for 17 to 41% of cases depending on the ascertainment criteria. These can be divided into two major categories, syndromic, where ID and additional clinical features would enable the diagnosis of a known syndrome, such as Down, Rett, Williams or Angleman, to name a few, which are associated with other dysmorphic clinical features (Galasso, Lo-Castro et al. 2010). Non-syndromic ID on the other hand is a case in which ID is mainly the only clinical feature and diagnosis of a known syndrome cannot be established. Throughout this report term ID will focus on the non-syndromic type.

1.1.1.1 Genetic causes of ID

Genetic factors that cause ID can either be due to *de novo* genetic variation (one that is not present in the parents) or an inherited mutation, with the parent either having ID manifested as a phenotype or with a parent being an unaffected carrier, if the disease is recessive, X-linked or not fully penetrant. The majority of genetic causes are chromosomal rearrangements and structural variations, such as copy number variants, and an increasing number of point mutations are being identified (Ellison, Rosenfeld et al. 2012; Mefford, Batshaw et al. 2012).

Genetic causes of ID vary in type and mode of inheritance. However, a large number of causative genes could lead to the same phenotype, and over half of the genes identified are genes involved in synaptic formation, signaling and plasticity. Understanding the role of these genes can aid in therapeutic intervention at early stages or even in the enhancement of memory function, which is one of the causative components for a low IQ (Verpelli and Sala 2012). For Rubinstein-Taybi syndrome (RTS) for example, two potential drug treatments have been tested on mouse models and they include suberoylanilide hydroxamic acid (SAHA), a histone deacetylases inhibitor that increases the level of late phase long term potentiation, and the other drug is Trichostatin A, which improved memory deficiency in CREBBP transgenic mice (Gropman and Batshaw 2010)

Besides structural variation there are epigenetic factors that could lead to deregulation by gene silencing that affects transcription (Galasso, Lo-Castro et al. 2010). This is mainly observed in imprinted disorder where the methylation pattern that causes gene silencing affects the phenotype, and in the case of Prader Willi and Angleman syndrome, the parental origin of the same methylated region leads to a different phenotype. The majority of epigenetic variations associated with ID are thought to involve a loss of function mutations in single genes. This suggests a potential therapeutic benefit through further understanding of the affected pathway with

potential design of drugs that can replace the function of the affected gene, such as the mentioned potential treatments earlier for RTS (Franklin and Mansuy 2011). Epigenetic causes are not the main focus of this study and so are mentioned briefly as another alternative to a possible cause.

1.1.2 Syndromic and Non-syndromic ID

As mentioned earlier, ID can be classified into two major groups: syndromic and non-syndromic. For the former it contains multiple clinical features and co-morbidities such as congenital heart malformation and can be diagnosed as a known syndrome. Non-syndromic ID (also referred to as idiopathic as the cause is not always understood), is mainly considered where ID is the only clinical feature or accompanied my milder features that are not associated with a known syndrome (Kaufman, Ayub et al. 2010)

1.1 Heritability and Genetic variation

1.1.1 Structural genetic variation

Structural variation is defined as deletions, duplications (tandem and interspersed), insertions, inversions and translocations that are greater than 1kb (Malhotra and Sebat 2012). These are illustrated in **Figure 1** below.

Genetic variation ranges in size from microscopically visible cytogenetic aberrations, which include aneuploidies, deletions, inversions and rearrangements, that are attributed to around 15% of ID cases, smaller micro-deletions or duplications, known as copy number variations (CNVs), accounting for another 15% of ID cases, while 10% could be attributed to point mutations and small insertions or deletions (indels), which

could be found in over 90 X-linked ID forms. This is in addition to a number of mutations on autosomal genes and epigenetic defects such as methylation that can lead to ID (Topper, Ober et al. 2011).

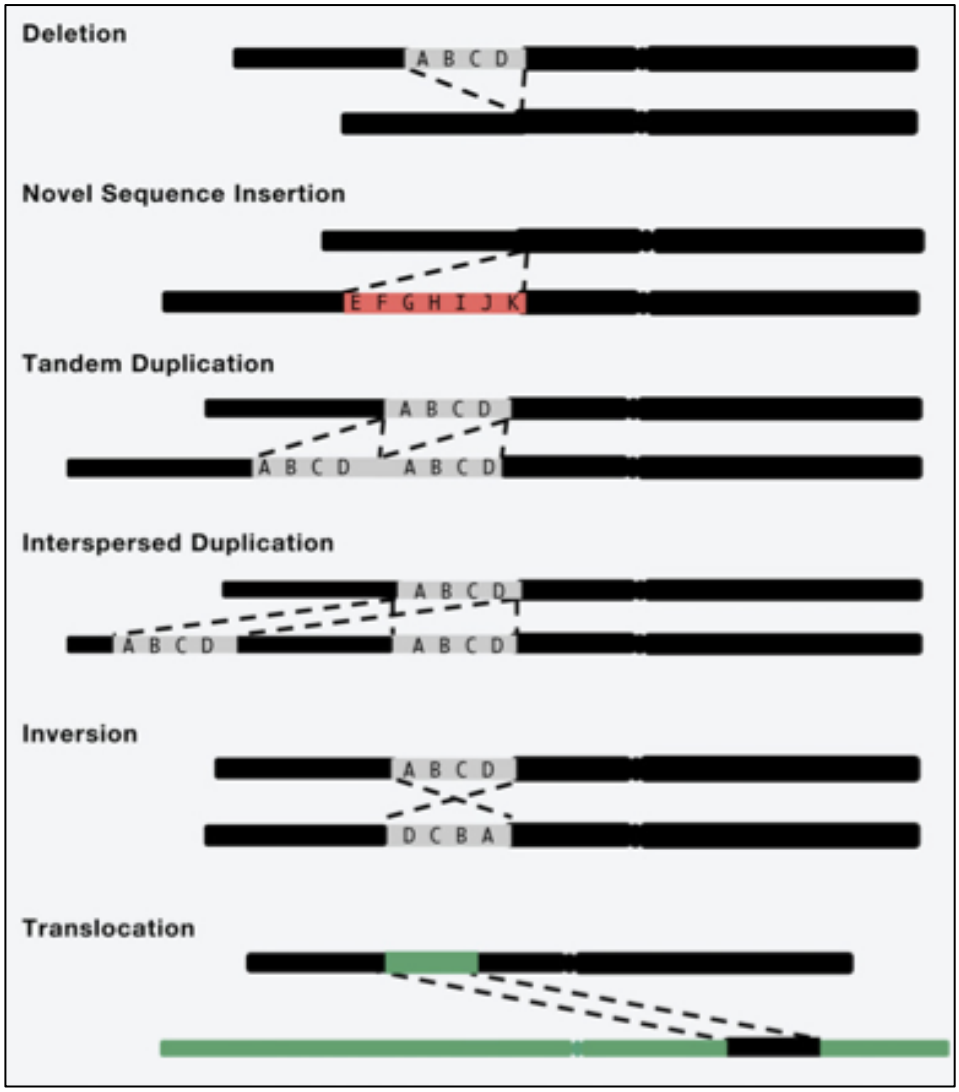


Figure 1: The different types of structural variations (Malhotra and Sebat 2012)

There seems to be a direct correlation between the size of the variation and the frequency of it found in the genome as illustrated in the graph in **Figure 2** below, with an increased risk of larger variations to be causative and

attributed to the disorder with increased severity. This is expected as larger variations could include more genes and hence has a larger impact on biological function (Girirajan, Rosenfeld et al. 2012).

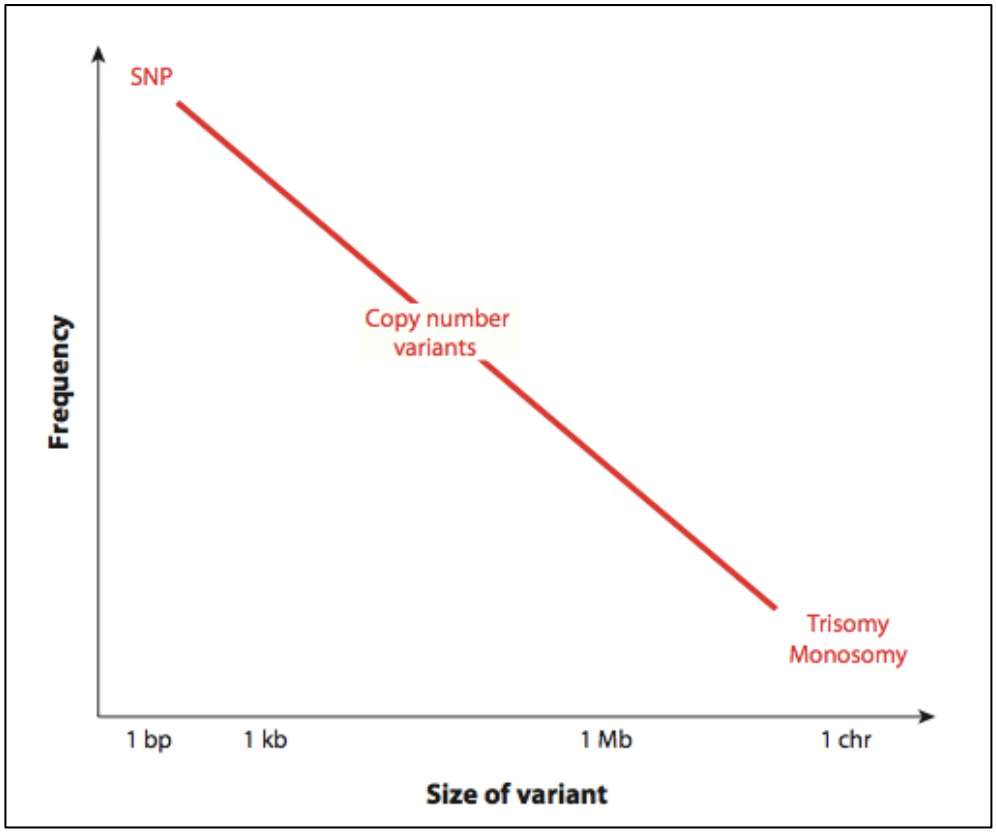


Figure 2: The correlation between size of variant and its frequency (Girirajan, Rosenfeld et al 2012)

1.2.1.1 Chromosomal

Chromosomal abnormalities are classical variants such as chromosome translocations, very large deletions or duplications or inversions that are typically visible under a light microscope using Geimsa staining or G-banding. 15% of ID cases have been attributed to chromosomal abnormalities. The most common is trisomy 21 (Down syndrome),

which is part of a numerical abnormality mainly resulting in the presence of an additional chromosome 21, accounting for around two thirds of cases. Other ID-related chromosomal abnormalities include deletions, unbalanced translocations and supernumerary such as trisomies or monosomies. However, cytogenetically visible abnormalities through traditional chromosome analysis methods are limited to a resolution of 5 Mb. This restricts the detection of smaller structural aberrations pushing for the use of alternative methods such as fluorescence in situ hybridization (FISH), which has a resolution of 1Mb, and further advanced techniques that include microarray analysis (Kaufman, Ayub et al. 2010; Crotwell and Hoyme 2012; Ellison, Rosenfeld et al. 2012).

1.2.1.2 Copy number variations

Copy number variations (CNVs) are defined as a copy number deletion or duplication that is bigger than about 0.5-1kb in size. It is estimated that 15 to 20 % of ID cases are thought to be due to submicroscopic CNVs (Galasso, Lo-Castro et al. 2010). However, not all CNVs are pathogenic. It is possible to detect more than 1000 CNVs in a normal individual. Rare CNVs are associated with pathogenicity and the likelihood is increased when it overlaps with regions of known genes associated with ID or a known syndrome with similar features. Some microdeletion syndromes that have been reported include, 1p36 microdeletion syndrome (Monosomy 1p36), 2q23.1 microdeletion syndrome, 2q37 deletion syndrome (Williams syndrome), 7q11.3 microduplication syndrome, 15q11-q13 microdeletion syndrome (Prader-Willi and Angelman syndrome), 16p11.2 microdeletion syndrome, and 17q21.31 deletion syndrome, (Galasso, Lo-Castro et al. 2010).

1.2.1.2.1 CNV formation

CNVs are thought to originate from two main pathways either as a result of unequal meiotic recombination or replication errors, often due to replication stress and can be induced by stress agents (Arlt, Wilson et al. 2012). The resulting CNV can be categorized into two main types; recurrent and non-recurrent CNVs. The former referring to regions that are commonly inherited and of fixed size while the latter refers to rearrangements of variable size and breakpoints (van Binsbergen 2011).

Four mechanisms are believed to lead to CNV formation and they include; nonallelic homologous recombination (NAHR), nonhomologous end-joining (NGEJ), Fork stalling and template switching (FoSTeS), and Retroposition. These are further illustrated in the figures below.

The nonallelic homologous recombination (NAHR) mechanism, as illustrated by **Figure 3**, occurs by the alignment of low copy repeats or any homologous regions and the consequential unequal cross over and recombination that would lead to duplication and deletion, these will segregate in the following cell division leading to a gain or loss in daughter cells.

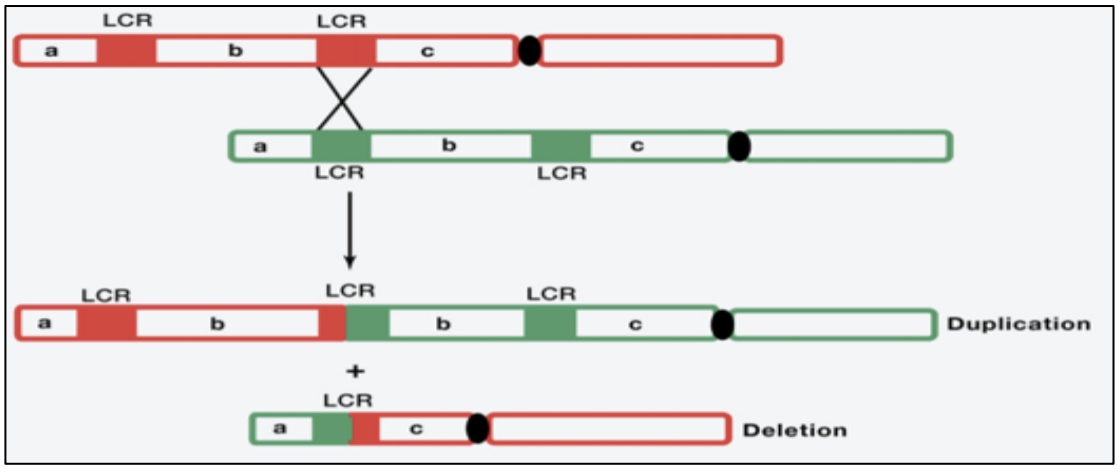


Figure 3: Nonallelic homologous recombination mechanism of CNV formation (Malhotra and Sebat 2011)

Non-homologous end-joining (NGEJ) is usually a product of DNA repair mechanism in which breaks in double stranded DNA at different stages of enzymic repair activities would lead to different possible outcomes based on the accuracy of the repair, illustrated by **Figure 4**. Products could include an accurately repaired fragment, deletion, microhomology mediated deletion or insertion.

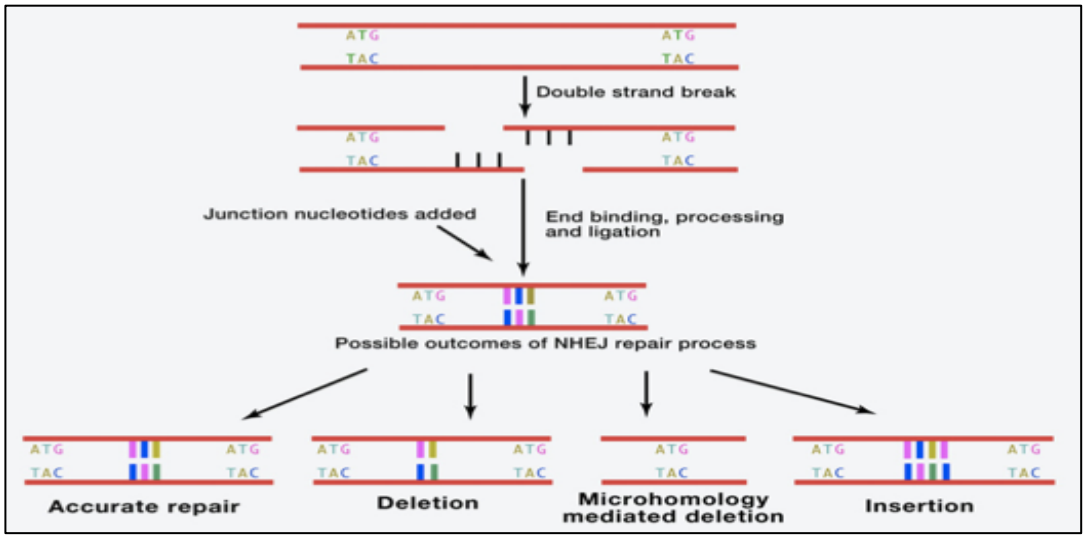


Figure 4: Nonhomologous end-joining mechanism of CNV formation (Malhotra and Sebat 2011)

The third mechanism is Fork stalling and template switching (FoSTeS), illustrated in **Figure 5**, this is due to replication errors where, a single strand break would stall the replication fork and then restarting at a different 3' end inversion leading to a duplication and deletion.

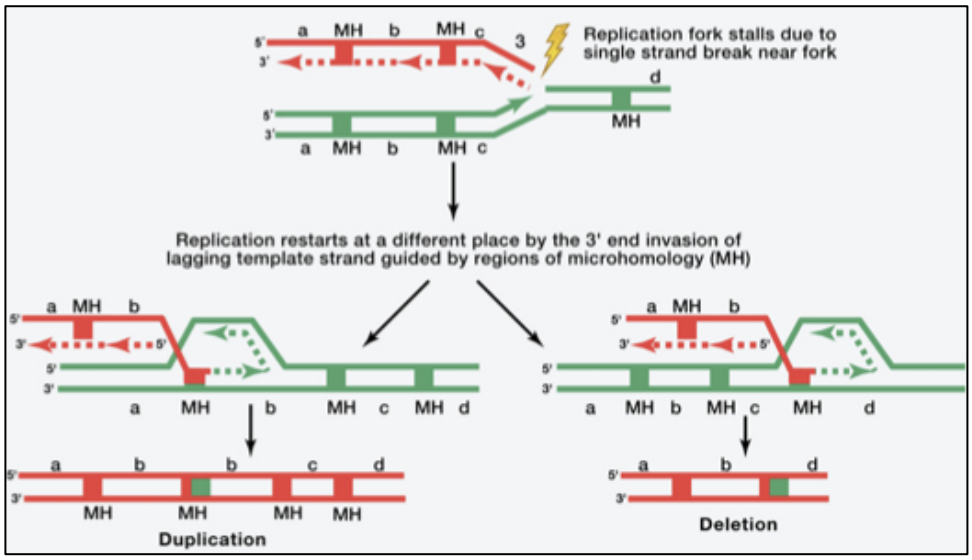


Figure 5: Fork stalling and template switching mechanism of CNV formation (Malhotra and Sebat 2011)

The fourth possible mechanism, illustrated in **Figure 6**, is based on retroposition by transposable elements known as long interspersed nuclear elements 1 (LINE1 or L1) that occupy around 20% of the genome and are present at around 500,000 copies present, with only 80 to 100 being active at full-length of 6kb.

As a result of the 1000 Genomes project, it was found that around 70% of deletions where considered to be the product of NHEJ, around 90% of insertions being the product of retroposition. Tandem duplications more likely caused by FoSTeS, while large deletions and duplications showed over 95% sequence identity at breakpoint suggestive of an NAHR mechanism (Malhotra and Sebat 2012).

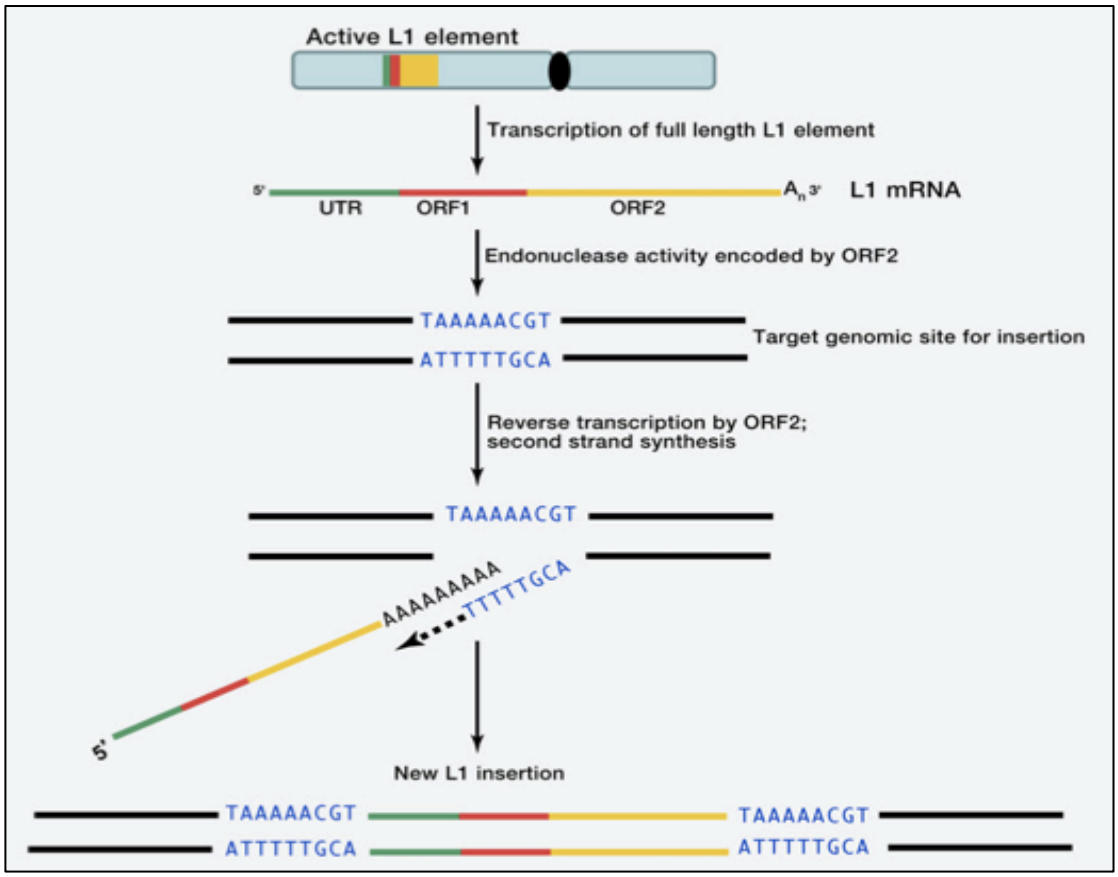


Figure 6: Retrotransposition mechanism of CNV formation (Malhotra and Sebat 2011)

1.2.1.2.2 CNV and ID

Several common CNVs have been curated in databases available for public access, these include the Toronto based DGV site (database of genomic variation) that catalogues common CNVs and hence comparison to this database can give an indication of how rare a CNV is (van Binsbergen 2011).

The literature is flooded with CNV suspected to be associated with ID, with some most likely being benign and others having been added as candidate regions as they cover regions overlapping genes linked with ID but depending on the identification method employed confirmatory test need to be applied before a conclusion about the pathogenicity is reached.

Few points worth noting about CNVs in ID, these include the heterogeneity o CNVs and related structural anomalies with many of them being private mutations or variations that are some times de novo or unique to the patient which makes it difficult to follow main criteria has centred on the size of the CNV, de novo status, and rarity of findings when compared to normal population.

Some recurrent CNVs include genomic segments involving 1q21.1, 1q41-42, 2p15-q16.1, 3q29, 7q11.23, 9q22.3, 12q14, 14q112, 15q13.3, 15q24, 16p11.2, 16p11.2-12.2, 16p13.1, 17p11.2, 17q21.31, 19q13.11, 22q11.2. These CNVs may be associated with specific syndromic features. Some recurrent CNVs appear to be inherited and/or incompletely penetrant. In a study of 1,227 individuals using microarrays it was found that the burden of CNV positively correlate to severity of disability. A published of potentially causative CNVs have been published and is the subject of the next section

(Morrow 2010; Girirajan, Brkanac et al. 2011; Girirajan, Rosenfeld et al. 2012; Grayton, Fernandes et al. 2012)

1.2.1.2.3 CNV morbidity map of developmental delay

In 2011 Cooper and colleagues published a list of CNVs associated with developmental delay after comparing CNVs in 15,767 children with intellectual disability and various congenital defects (cases) to CNVs in 8,329 controls. Although the phenotypic features are mixed in the cases, nevertheless the large number of samples used helped to eliminate common CNVs and further understanding of their potential pathogenesis. Rare CNVs were found at a frequency of less than 1% in the population are more likely to be pathogenic (Cooper, Coe et al. 2011).

It was observed that 25.7% of cases had CNVs of > 400Kb compared to 11.5% of the control with greater difference as the CNV size increases so for example 1.5 Mb sized

CNVs were found in 11.3% of cases compared to 0.6% of controls. In addition, CNV burden varied based on clinical features with more large CNVs found in craniofacial abnormalities and heart defects as opposed to epilepsy and autism spectrum disorder. This list of potentially pathogenic CNVs was constructed to aid researchers in identifying causative CNVs taking advantage of the large cohort studied (Cooper, Coe et al. 2011).

1.2.1.3 Point mutations

This does not fall under the general umbrella of structural variation but rather as sequence variation. Point mutations or single base substitutions, that causes the replacement of single nucleotide with another that disrupts gene expression, has been reported in many studies. These can be detected with methods for single nucleotide polymorphism (SNP) or variation (SNV) and a number of these have been reported to cause ID (Rizzi, Beunders et al. 2012). With advances in next generation sequencing, more of these sequence variations have been identified and a replacement that can cause malfunction in any of the over 300 genes associated with ID would be considered pathogenic (Rauch, Wieczorek et al. 2012; de Ligt, Boone et al. 2013).

To name a few, Harvey and colleagues reported sequence variations in the MECP2 gene in patients with ID, including a point mutation of c.49G>A that causes an amino acid change (Harvey, Menon et al. 2007) and point mutations were reported in gene ZNF407 in patients with moderate non-syndromic intellectual disability and autism (Ren, Liang et al. 2013).

1.2.2 Modes of inheritance

1.2.2.1 Mendelian

1.2.2.1.1 X-linked

The most common genes that cause ID are located on the X chromosome, with Fragile-X and Rett syndrome being the most common forms of X-linked ID. Other genes mutations such as those found in NLGN3 and NLGN4 have been also associated with ID (Verpelli and Sala 2012). Accounting for 5 – 10 % of ID cases, perhaps it is the most studied form of ID, with many reported point mutations (Lubs, Stevenson et al. 2012)

1.2.2.1.2 Autosomal dominant

Several autosomal dominant mutations have been linked to ID such as 10 de novo mutations identified in genes known to be associated with ID as reported by de Ligt and colleagues via exome sequence analysis of a cohort of 100 patients with severe ID (de Ligt, Boone et al. 2013). In the same study no causative autosomal recessive mutations were observed, which supports a hypothesis that the majority of severe ID is caused by genetic de novo mutations, which are not present in the parents. This comes at no surprise as mutations in any of over 1000 different genes could lead to ID.

One method of detecting genes involved in autosomal dominant is by breakpoint analysis of chromosomal aberrations. Information about breakpoint locations and how genes are disrupted lead to the identification of genes such as DOCK8, MBDS, CDH15 and KIRREL3 (Kaufman, Ayub et al. 2010)

1.2.2.1.3 Autosomal recessive

There are a number of reported autosomal recessive variations and mutations, where by two copies of aberration or variation are required for the disease to lead to a phenotype under the classical Mendelian model.

Autosomal recessive ID is the most common type of ID in populations with frequent parental consanguinity and it is also not uncommon in outbred Western populations (Ropers 2008; Ropers, Derivery et al. 2011; Musante and Ropers 2014)

Usually, this form of inheritance comes from familial forms of ID where parents appear normal and the chances of reoccurrence would be higher than the de novo counterpart. These recessive variant have been rare in outbred populations yet successfully identified in large consanguineous families using a method known as homozygosity mapping (Kaufman, Ayub et al. 2010).

Autozygosity or homozygosity by descent refers to the inheritance of the same variant from a common ancestor through maternal and paternal linages. Children of consanguineous marriages have more homozygous regions (originating from a common ancestor) than their outbred counterparts, which increase the probability of passing down rare recessive variants within these regions (Kaufman, Ayub et al. 2010). It is estimated that there is 10 times more recessively inherited congenital conditions in consanguineous populations than outbred ones and that these often result in early death.

Homozygosity mapping has been widely used in identifying regions that harbor pathogenic recessive variants due to the likelihood of the cosegregation of the homozygosity region with the pathogenic variant in the affected individuals. It has been utilized as a very useful diagnostic tool and recommended option for consanguineous families (Alkuraya 2010). More on homozygosity and consanguinity is covered in chapters 2 and 6.

As such a homozygous region on chromosome 11p have been identifies in two independent studies one through mapping of an extended 5 generation Pakistani family with autosomal recessive severe non-syndromic ID in which a 6MB homozygous block in the telomeric region of 11p15 was identified to cosegregate with the phenotype (Rehman, Baig et al. 2011), while in a study of consanguineous Syrian family a 6.4MB homozygous region was reported amongst ten other regions and a mutation was identified in TRAPPC9 gene within a homozygous block located on 8q24 (Abou Jamra, Wohlfart et al. 2011).

1.2.2.2 Non-Mendelian inheritance

Non-Mendelian inheritance is often associated with modes of inheritance that do not follow the classically expected models of recessive, dominant or X-linked inheritance. In some cases of ID there is another underpinning mode that is evident in its varied penetrance exemplified by the heterogenous phenotype whereby affected siblings would display different severity of ID. For these cases an oligogenic mode of inheritance is expected whereby the effect of a number of genes would lead to disease manifestation and could be dosage sensitive. Another suspected model is that of a multifactorial nature whereby early exposure to environmental agents would lead to the triggering of susceptibility genes or an epigenetic switching that would lead to the disorder, this model is sometimes applied to autism, where ID is a clinical feature in some but not all cases.

1.2.3 Common variant vs. rare variant model

Two models exist that form the basis of the method of screening and variant identification they are based on the hypothesis of common variant causing a common disease or a rare variant causing a common disease. The latter is the likely model when

it comes to heterogenous ID in the general population (Heil and Schaaf 2013). This makes it difficult to pool samples to identify a common cause as a variable one is suspected and hence large family studies, where the underlying genetic cause would be shared, makes them ideal for tracking the rare variant.

1.2.4 Phenotypic heterogeneity

As discussed earlier ID diagnosis includes a feature of having an IQ score less than 70 and this can be further classified based on the IQ score to range from borderline, mild, moderate, severe and profound IQ. This adds to the phenotypic mix of ID and it is worth mentioning that the ID score is based on a multiple component tests that yield to the overall score such as tests on memory and ordering objects. Hence, the underlying genetic cause could differ from case to case, although the general phenotype is that of ID differing only in severity. Conversely phenotypic heterogeneity has been observed were a deletion in the same region has lead to a different phenotype.

Such is the case of chromosome 15q13.3 deletion that has been associated to ID, schizophrenia, autism and 1% of idiopathic generalized epilepsy. These phenotypic discrepancies have been difficult to explain and Girirajan and colleagues have proposed a “two-hit” model based on a common observation where affected individuals of a 16p12.1 chromosomal microdeletion have additional CNVs compared with controls. This support the oligogenic mode of inheritance in which small number of rare variants of a large effect causes a heterogeneous genomic disorder (Girirajan, Rosenfeld et al. 2012). Another observation in their study was that males were more sensitive to effects of large variants than females in X-linked disorders, which is mainly attributed to the presence of a single copy of the X chromosome.

1.2 Rubinstein-Taybi Syndrome

1.3.1 Background

Rubinstein-Taybi Syndrome (RTS, OMIM 180 849) is a rare congenital syndrome with an incidence of 1 in 100,000 to 125,000 newborns (Roelfsema, White et al. 2005), affecting males and females equally (Hallam and Bourtchouladze 2006). This autosomal dominant disorder is characterised by growth and mental retardation, skeletal abnormalities, with typical facial features, broad thumbs and halluces (big toes) (Bartsch, Schmidt et al. 2005; Bartholdi, Roelfsema et al. 2007). Other associated features include predisposition to tumours and keloid (area of irregular fibrous tissue at the site of injury or scar) formation (Bartsch, Locher et al. 2002).

The first report of Rubinstein-Taybi syndrome (RTS) was initially made by Michail and colleagues in 1957 under the name of *broad thumb-hallux syndrome* and was later described by Rubinstein and Taybi in 1963 (Hsiung 2004), where seven children with mental handicap had the characteristic broad thumbs, big toes and very similar facial features. RTS was estimated to account for one case per 300 institutionalized subjects with IQ estimates ranging from 20 to 80 (Berry, Smith et al. 1987).

This review will cover the genetic causes as reported by different studies, the phenotype and genotype correlation.

1.3.2 Genetic causes

Petrij and colleagues first reported the disease-causing gene to be localised at the chromosomal region of 16p13.3 where the CREBBP gene is located (Petrij, Giles et al. 1995). This was found using fluorescent in situ hybridisation (FISH) technology that mapped microdeletions of between 130 and >650Kb in RTS patients to this region

(Petrij, Giles et al. 1995). The majority of translocations reported were within this region and inclusive of the CREBBP genes.

Gene abnormality can be detected through various methods including FISH, real-time PCR, direct gene sequencing, as well as, the use of Multiple Ligation-dependent Probe Amplification (MLPA) to detect CREBBP deletions. Of which MLPA seems to be a rapid and cost effective technique for CREBBP mutation detection, although MLPA would not allow detection of mosaic cases whereas FISH will (Roelfsema, White et al. 2005).

The CREBBP gene (Also known as CBP) is 150kb long and encodes a 2442 amino acid sequence (Bartsch, Locher et al. 2002). Several mutations in the CREBBP gene have been reported by different groups, which are mainly *de novo* and with no particular mutation hotspots. Bartsch and colleagues reported 11 truncation mutations and one missense mutation. Whereas, Coupry et al, reported 27 mutations identified in 63 RTS patients of which 10% had cytogenetic deletions of band 16p13.3 and 47.6% had CREBBP gene abnormality including microdeletion, duplication, or point mutation (Coupry, Roudaut et al. 2002).

For 10% of RTS cases, large cytogenetic deletions or translocations at the 16p13.3 band have been observed, which cover the CREBBP area. Whereas other mutations have been found within the CREBBP gene. A detection rate of 56% of RTS cases with CREBBP mutations amongst RTS patients was reported, as detected by direct sequencing. Several polymorphisms were also identified along the gene sequence (Bartsch, Schmidt et al. 2005). A similar detection rate of 57% was also observed in 60 RTS patients, using denaturing high performance liquid chromatography, to detect mutations in the CREBBP gene (Udaka, Samejima et al. 2005).

Moreover, five deletions in the CREBBP gene and 14 mutations (10 truncations, one splice site, three new missense mutations) were reported among 31 Italian RTS patients, yielding a 61.3% mutation detection rate in CREBBP gene (Bentivegna, Milani

et al. 2006). This higher rate of detection is achieved by combining different detection methods to screen for mutations in the CREBBP gene.

In CREBBP most mutations occur de novo with no significant hot spots although mutations within the Hat domain seem to be mainly causative for RTS. However, the *CREBBP* gene harbors an unstable region around exon 2. This region was is thought to be unstable because all translocation and inversion breakpoints could be found in most patients with RTS, as well as leukemia breakpoints for which CREBBP functions as a fusion partner. This region has a higher percentage of interspersed repetitive elements, which makes it more susceptible to breakage and recombination (Gervasini, Castronovo et al. 2007).

There are two proposed models of how mutations in CREBBP may lead to RTS. The first model being that of haploinsufficiency, whereby two functional copies of the normal, dosage sensitive, gene is required to produce enough product to sustain normal function and development. The second is the dominant negative effect, in which, abnormal product derived from the mutant allele may inhibit the normal wild-type product (Hallam and Bourtchouladze 2006). The dominant negative inhibiting mechanism is thought to be at play, as supported by in vitro studies in which microinjections of CREBBP domain would block the normal activation of the reporter gene (Oike, Takakura et al. 1999). Deletions of the CREBBP gene results in RTS, which supports the haploinsufficiency model and the dominant negative mechanism was not proven in humans.

Another RTS causing gene is EP300, which is located on the chromosome 22q13.2 position with mutations detected in 3.3% of RTS patients (Roelfsema, White et al. 2005). EP300 mutations are ten fold less frequent than CREBBP mutations but this could be due to the selection criteria of RTS patients.

Roelfsema and colleagues searched for mutations in the EP300 gene as it shares homology with CREBBP. Both include a HAT domain, which is thought to be essential

for their function. Most CREBBP mutations in RTS cases seem to cluster in the HAT domain although few mutations have been found out side of the HAT domain.

Some cases with the EP300 mutations do not have all of the classical features of RTS (Bartholdi, Roelfsema et al. 2007). Three RTS cases and a fourth that does not have classical RTS phenotype where found to have mutations in the EP300 gene as reported by Roelfsema and Zimmermann respectively (Zimmermann, Acosta et al. 2007).

Other genes have been detected along side the CREBBP region to include DNASEI and TRAP1 (Bartsch, Rasi et al. 2006). Mutation in the former gene has been associated with a disorder of the autoimmune system, which might be associated with a high rate of infection.

Besides microdeletions, duplications have also been observed in other cases. Thienpont and colleagues reported a case, with mental retardation and congenital heart defect and no RTS feature, to have a microduplication spanning TRAP1 & CREBBP genes and part of ADCY9 and DNASE1 (Thienpont, Breckpot et al. 2007).

Likewise, Marangi and colleagues reported another case with a 1.7Mb duplication at the region encompassing the CREBBP gene that is thought to be causative of the multiple congenital anomalies and intellectual disabilities (MCA/ID) syndrome observed (Marangi, Leuzzi et al. 2008).

The low rate of detection still points out at the possibility of other genes being involved in RTS and so using array CGH analysis Gervasini et al found deletions and duplications in seven cases out of 26 who were previously found to be negative for any CREBBP mutations. These deletions/duplications ranged from 0.5 to 9.1 Mb in size and found on different chromosomes in different patients with three of them having such copy number variation on chromosome 2. The other regions are located on chromosomes 3, 7, 17 and 18 (Marangi, Leuzzi et al. 2008).

This further supports the genetic heterogeneity of RTS and mutations or deletions in other gene still need to be confirmed as causative by other methods.

Another possible cause of missed CREBBP mutation detection could be due to mosaicism, which can only be detected by few techniques such as FISH. Cases with mosaicism have been reported in few studies, Hennekam et al 1990 and Stevens et al 1990, as reviewed by Chiang. Also Gervasini and colleagues identified 3 mosaic CREBBP mutation through FISH and microsatellite analysis of 42 Italian RTS cases (Gervasini, Castronovo et al. 2007).

Chiang also reported two families one with two affected sibs with the same mutation and hence a suspected germline mosaicism that still requires confirmation from sperm DNA as no mutation was detected by both blood and saliva DNA. However the father had broad big toes, a feature of RTS. In the second unrelated family the father also had the same broad big toe feature and low-level mosaicism was confirmed in the blood DNA through bidirectional DNA sequencing (Chiang, Lee et al. 2009).

1.3.3 CREBBP & EP300 functions

The CREBBP gene (also referred to as CBP) encodes the CREB-binding protein. It is involved in basic cellular functions including growth, differentiation, DNA repair and apoptosis; it also regulates the expression of many genes (Bentivegna, Milani et al. 2006).

CREBBP codes for a nuclear phosphoprotein, which acts as a coactivator in cyclic-AMP-regulated gene expression. It promotes gene transcription by binding to the transcription factor CREB (Schorry, Keddache et al. 2008).

In neurons, CREBBP is activated in various processes such as learning and memory via protein kinases (PKA), calcium/calmodulin-dependent kinase IV(CaMKIV), and mitogen-activated protein kinase (MAPK), which might indicate the deficit of cognitive ability might be attributed to the loss of long term memory formation in RTS patients (Hallam and Bourtchouladze 2006).

On the other hand, EP300 codes for the ubiquitous nuclear phosphoprotein p300 which also acts as a transcriptional coactivator. There is a wide involvement of p300 in cell-type and signal specific manner of many gene expressions and hence its role in cellular proliferation, apoptosis, and embryogenesis. Also, p300 plays a significant role in epigenetic regulation through catalysing acetylation of histones and transcription factors. Abnormalities in p300 function is linked to deregulated target gene expression (Ghosh and Varga 2007).

Although CREBBP shares a high degree of homology with EP300, they nevertheless, do not completely overlap in function (Bartholdi, Roelfsema et al. 2007). Both are global transcriptional coactivators. They possess conserved domains including a histone acetyltransferase (HAT) domain involved in DNA unfolding during transcription (Bartsch, Schmidt et al. 2005).

It has been shown that the loss of the HAT activity is sufficient to cause RTS (Bentivegna, Milani et al. 2006). In addition, there are reported links between HAT activity and long-term memory (Roelfsema and Peters 2007), which may account for mental retardation.

1.3.4 Phenotype to genotype correlation

There is some controversy on whether RTS patients have minor variations of the same symptoms or that the disorder includes a spectrum of very mild and severe cases (Bartsch, Rasi et al. 2006). On the whole, phenotypically patients with CREBBP and

EP300 mutations are similar as reviewed by Roelfsma (Roelfsema, White et al. 2005). Yet, features seem to range between cases in description, from being incomplete or RTS-like to severe RTS cases.

For instance, mild missense CREBBP mutation cause incomplete RTS where patients had RTS features but normal intelligence, less typical facial features, and/or atypical skeletal changes than classical RTS (Bartsch, Rasi et al. 2006).

Conversely, Bartsch described severe RTS, which the authors also refer to as a possible separate syndrome 16p13.3 deletion syndrome. They reported a higher rate of infection and mortality amongst these individuals than classical RTS. These individuals had a deletion of the DNASE1 gene, which lays around 60 bp close to he CREBBP gene.

Stef and colleagues stated that there was no phenotypic difference between RTS cases with small or large deletions in the CREBBP gene except for one case where the whole gene and an additional large centromeric region in which case the patient had a severe phenotype and died aged 34 days (Stef, Simon et al. 2007). Two other patients in the same study had a deletion that included the DNASEI gene and have a severe form of the disease, however, they did not have the high susceptibility to infection nor severe mental retardation, which were previously reported by Bartsch 2006. Also, in terms of severity, duplications have been suggested to cause less severe phenotype than deletions (Thienpont, Breckpot et al. 2007).

It is worth noting that there are secondary features that are observed in a proportion of cases and not all such as heart malformations and kidney problems or oro-dental features. However, the main clinical features that would lead to a diagnosis of RTS are the typical facial features (including microcephaly, down-slanting eyelids, beaked nose and arched palate) and the broad hands and big toes. Even the last is sometimes not found in some cases which coincidently where shown to have mutations in the EP300 gene (Roelfsema and Peters 2007).

Although the number of RTS cases with mutations in EP300 remains small, Bartholdi and colleagues described the reported patients to have normal hands and feet and thus recommended that RTS diagnosis criteria must be expanded to include patients without broad thumbs and halluces (which are generally noted as the main recognizable features of RTS) and so more patients might be diagnosed (Bartholdi, Roelfsema et al. 2007). Two more patients with E300 mutations and different facial features from classic RTS have been reported (Bartsch, Kress et al. 2010). Zimmermann et al suggest that EP300 mutations could be found in cases with non-classical RTS. The mutation (1bp deletion) they reported was outside of the HAT domain, which could explain the non-classical features of their case (Marangi, Leuzzi et al. 2008).

In some cases, duplication or deletion within the same gene can yield to different phenotypes as observed with the PMP22 gene, where a duplication leads to Charcot-Marie-Tooth neuropathy while a deletion may lead to a mild or asymptomatic HNPP, other such examples are also mentioned in the paper by Marangi and colleagues (Marangi, Leuzzi et al. 2008). Duplications have been suggested to cause less severe phenotype than deletions (Thienpont, Breckpot et al. 2007).

Furthermore, Schorry and colleagues sequenced the DNA of 93 patients and reported no significant difference between the genotype and phenotype except for growth retardation, seizure disorder and autistic features (Schorry, Keddache et al. 2008).

Lastly, studies on animals revealed different contributions of EP300 and CREBBP genes to RTS etiology as demonstrated by mice with EP300 gene mutations having milder cognitive impairment than those reported with CREBBP gene mutations (Viosca, Lopez-Atalaya et al. 2010).

1.3.5 RTS to date

Around 60% of RTS cases where caused by anomalies in the CREBBP gene. With EP300 gene mutations accounting for an additional 3% of cases leaving a large percentage of cases with an unknown genetic cause. This could be in part due to mosaicism, incomplete coverage of the CREBBP and EP300 gene sequencing or due to additional genes that could be involved.

The same translocation t(2;16)(q36.3;p13.3) have been reported in different RTS patients by different groups with the fourth case reported by Torres et al 2010. Interestingly, Gervasini recently reported a duplication at 2q34-35 amongst other copy number variations in RTS cases as well as two deletions on chromosome 2, (2)(q22.3q23.1) and (2)(q24.3q31.1). These deletions include few genes of interest such as the zinc finger homeobox 1B (ZEB2), which is associated with Mowat-Wilson syndrome (OMIM 235730), HOXD13, which encodes a transcription factor involved in morphogenesis, DLX1 and DLX2 genes that could be responsible for craniofacial patterning in addition to differentiation and survival of inhibitory neurons in the forebrain. Another gene, which lies within one of these deleted regions, is the ATF2 gene (activating transcription factor 2 also known as CREBP1 or CREB2), which encodes a product that stimulates CRE dependent genes and involved in HAT activity for H2B and H4 (Gervasini, Mottadelli et al. 2010).

Other genes that were found in other genomic imbalances include HDAC9 and TWIST1 in a patient who also has features of Saethre-Chotzen syndrome (SCS; OMIM 101400), in which TWIST1 gene is thought to be involved (Gervasini, Mottadelli et al. 2010).

Other disorders share impaired CREBBP/CREB function, namely Huntington's disease, Alzheimer's disease, and amyotrophic lateral sclerosis (Hallam and Bourtchouladze 2006). How mutations lead to RTS specifically should be distinguished. Which again might suggest a more complex mechanism at play or a number of genes that synergistically lead to RTS.

It is worth noting that the difference in the mutation detection rate reported by different groups could be due to differences in inclusion criteria with some groups including RTS-like cases in there study as part of the wider RTS spectrum and others only cases with classic RTS clinical features. The low level mutation detection rate points at the genetic heterogeneity of RTS with possibly more genes at play that could be related to the CREBBP and EP300 pathways as well as possible epigenetic malfunctions since both CREBBP and EP300 products are involved in acetylation.

Gervasini supports the possibility of there being a spectrum of RTS as some do not manifest all of the reported clinical features, which may account for the possibility of variable genetic causes.

Moreover, screening for EP300 gene is usually done on cases that are negative for CREBBP mutations and so it is still remains to be determined if there are mutations within the cases that have CREBBP mutations. Also the possibility of cases of EP300 cases being of another syndrome that is overlapping with RTS in features. Three further cases with EP300 mutations have been reported (Foley, Bunyan et al. 2009; Bartsch, Kress et al. 2010) with milder features and slightly different that RTS, which could point out to a RTS like syndrome that could be distinct from classical RTS.

The involvement of p300 protein in embryogenesis might account for the low mutation rate observed as major deletion might affect viability. This is yet to be proven but could remain a possibility. Also, RTS cases with EP300 gene mutations might be part of a different disease all together and so within this distinct group EP300 mutation detection rate would be much higher if these cases are separated from classical RTS.

Furthermore, due to the diverse roles played by CREBBP and p300 proteins, gene analysis of RTS cases may aid in further understanding of transcription, cognition and brain function, keloid formation, and cancer (Bartsch, Locher et al. 2002).

It is very important to diagnose the cases properly and identify the underlying cause in order to the appropriate management. This could include possible future intervention especially for drugs that target epigenetic malfunctions.

Finally, additional family studies need to be carried out to determine the level of mosaicism, which could explain the cause for some cases and would shed light on the inheritance pattern of RTS and subsequent reoccurrence rate that would aid in counseling.

1.4 Current clinical diagnosis of ID in Kuwait

The main genetic diagnostic laboratory is based in the Kuwait Medical Genetic Centre (KMGC) with limited ties with research facilities and when these are established they are often limited to few disorders where there is research interest by reputable universities or genetic centres abroad. Nevertheless, such collaboration takes a long process to set up and in the majority of cases diagnostics is offered on locally available tests or is reviewed on case per case basis on whether a known genetic cause has been identified, in which sending samples abroad to a specialized centre for a specific test would be reviewed by a specialist committee, taking into consideration costs and number of test required as well as how informative the results would be and scope of benefit. For example, a disorder with a few known genes with multiple affected members would be considered for testing opposed to a rare disorder with unknown genetic cause in a single individual. In addition, the urgency of having the results is also considered, in cases of cancer for example, it would have an impact on advised treatment or intervention.

Within KMGC, the genetic diagnostic testing carried out on cases with ID and or congenital malformations relies heavily on cytogenetic karyotyping and in some cases FISH technique is then applied to identify known chromosomal abnormality. On the

molecular level only few tests are available for specific syndromes such as Fragile X expansion, and it is automatically offered as a test to any case of ID cases upon confirming a normal karyotype, with around 500 cases being tested per year.

Advances in the field have sparked ambitions to start investing in new machines and techniques that would further advance the genetic services on offer. Intellectual disability is one of the major disorders referred to KMGC and it is often to establish disability and report that it is from birth in order for parents to claim benefits offered by the government that varies depending on the condition and severity of the case, which can include accommodation subsidies, medical support and tuition fees for special need schooling.

For some families ID disorder is inherited leading to multiple affected members and so parents seek genetic counselling for further family planning, which in some cases can have a social consequence that could lead to a divorce or a second marriage, which is religiously allowed and permissible by law, and would be justified for some as a right for “healthy offspring”, in which the first wife would be expected to accept such an arrangement.

In addition, there is a high incidence of first or even double cousin marriages and cases with a family history of ID would be referred to the KMGC for a premarital genetic counselling session or for family planning of additional pregnancies. It is important for any session to have an estimated risk that is based on the genetic etiology of the disorder and taking into account the mode of inheritance and if any testing can be offered for predisposing elements. In the case of non-syndromic ID disorder there is no established set up for genetic testing beyond Fragile X FRAX-A and E expansion test. At the moment counselling sessions would end up with very vague information of the disorder after routine checks are made and found to be negative. If a close syndrome is suspected then samples can be sent abroad given that causative genes are known or that it will be part of a big research project with set collaborations such as autism

spectrum disorders. Remaining cases would be regarded as unsolved and families would be left to make decisions on the uncertain information given. Many would hope for IVF treatment, which would not be offered without knowing the causative mutation.

Based on the need to provide proper diagnosis and risk calculation it is vital to establish a screening protocol for idiopathic ID and rare syndrome gene detection. Hence, as a starting step the aims of my project are to:

- Collect DNA and pedigree information on families with RTS and multiple cases of ID from the Kuwaiti population.
- Detect known and novel CREBBP mutations and deletions in the Kuwaiti population by HAT domain sequencing and copy number analysis.
- Identify novel RTS causing gene(s) in cases without CREBBP mutations, using microarray based copy number variant analysis, and explore whether these mutations can explain other types of mental retardation.
- Use copy number variant analysis to detect deletions and duplications in consanguineous multiply affected families with intellectual disability using DNA microarray analysis and the Multiple Ligation Probe Assay.
- Evaluate the use of Cytogenetic arrays and MLPA techniques in CNV detection
- Follow up specific CNVs identified with gene expression and bioinformatic analysis.
- Provide recommendations for the establishment of an ID screening and gene detection protocol.

The aims were tested with the following underpinning hypotheses

Hypotheses

- *Hypothesis 1: Consanguineous families with multiple affected members harbour recessive mutations for ID syndrome.*
- *Hypothesis 2: There are known and novel CREBBP deletions and HAT domain mutations in the Kuwaiti population, which cause RTS and other ID phenotypes, which can be detected by MLPA and sequencing analysis.*
- *Hypothesis 3: There are novel or known copy number variants in other regions of the genome, which causes non-CREBBP forms of RTS or familial mental retardation. Some of the regions of interest will harbour recessive genes that can be further analysed by homozygosity mapping.*
- *Hypothesis 4: There are novel gene mutations in the genome causing recessive ID, which can be detected by exome sequencing in index cases.*

Chapter 2- Sample Collection

2.1 The Geography of Kuwait

Kuwait is a Middle Eastern country located in Southwest Asia, sharing borders with Southern Iraq, Northern Saudi Arabia and the Arabian Gulf on the East. This small state covers 17,820 square kilometers, around 200 kilometers from north to south and 170 kilometers across. Kuwait was the Centre of pearl diving and boat construction until the discovery of oil, which became the major export. Kuwait is split into six governorates, Al-Asimah (where the capital is located), Hawalli, Mubarak Al-Kabeer, Al Farwaniya, Al Ahmadi and Al Jahra (**Figure 7)**.

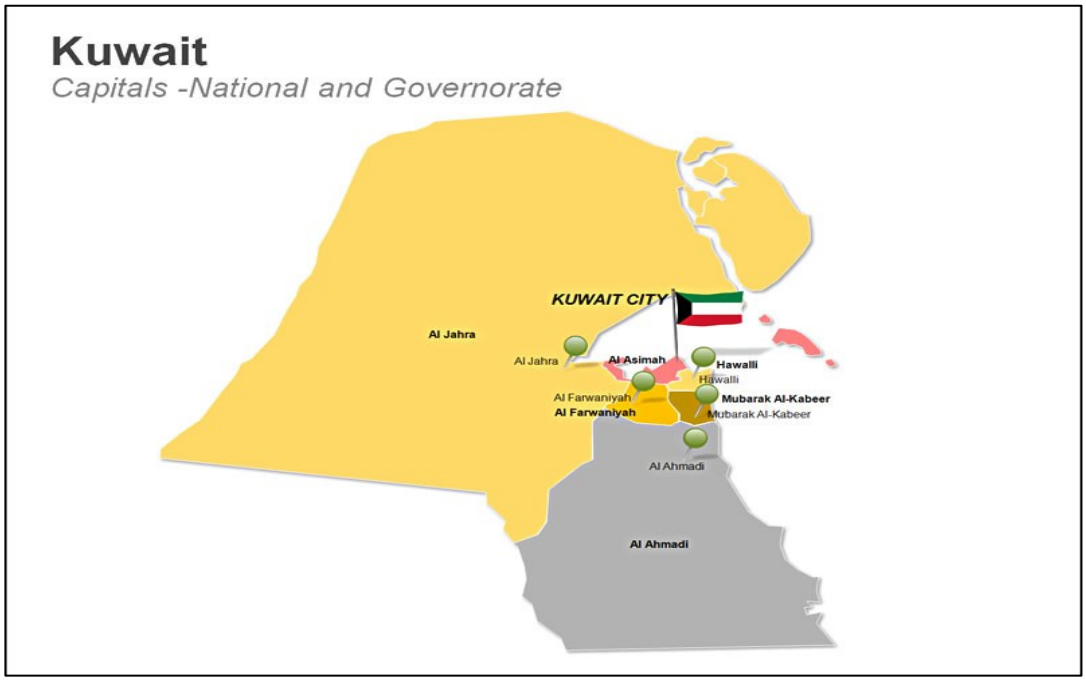


Figure 7: Map of Kuwait Governorates (from www.24point0.com)

2.2 The Demography of Kuwait

The Kuwaiti population consists of 96% urbanized and 4% nomadic or semi-nomadic people, reaching an estimated total of between 3 and 3.5 million including both locals and foreigners at a ratio of about 1: 2.

In terms of ethnicity, 60-65%, of the Kuwaiti population, including non-Kuwaitis, are of Arab descent; 30-35% is Asian (including South East and Far east Asians), while the remaining 5-10% consists of other nationals including mostly Africans, East Asians and Europeans.

Although Kuwaitis refer to themselves as Arabs, they have various origins with some originating from Persia and others from non-Arab countries such as sub-Saharan Africa, while some can trace themselves back to Arabic tribes from the Arabian Peninsula or Iraq. The majority of those living in Kuwait (85%) are Muslim and the remaining are of other religions including Christian and Hindu.

Literacy rate is at a 93% high, due to government support with free education, including University level, although not inclusive of foreign residents. Arabic is the official language of Kuwait, but most Kuwaitis are bilingual and speak English or Farsi and most foreigners speak English, Hindu, Urdu, Filipino or Bengali.

2.3 The Kuwait Medical Genetic Centre

2.3.1 Location and coverage

The Kuwait Medical Genetic Centre (KMGC) is located in the Al-Asimah governorate beside the maternity hospital and within the main medical district (see figure 2.2 for the location). Genetic counselling services first started in February 1979 with different genetic clinics at Al Sabah Hospital, Amiri Hospital, Al-Jahra Hospital, Addan Hospital,

Maternity Hospital, and in Kuwait Cancer Centre to offer free genetic counselling to the whole population of Kuwait and also some Gulf Cooperation Council (GCC) states.

The former minister of health Dr Abdul Rahman Al Awadi officially established the centre in 1980 with genetic services being offered in genetic clinics on weekly basis and analysis supported by the Biology Department in the Faculty of Science at Kuwait University. For molecular testing samples were sent to the US or the UK with focus on genetic disorders including; infertility, cases of repeated spontaneous abortion, congenital malformations, cases of suspected chromosomal disorders, familial ID, hearing impairment, ophthalmological disorders, short stature, and haematological diseases.

KMGC was rebuilt after the Iraqi invasion in 1990 and was developed to contain a molecular genetics laboratory, fluorescence *in situ* hybridization (FISH) laboratory as well as the cytogenetic laboratory. In 2009 the KMGC was moved to a newly built centre near the old centre and serves as the main centre for all the genetic clinics (**Figure 8**). KMGC will be referred to as the Genetic Centre in the rest of the thesis.

All genetic testing in the public sector is now carried out in the Genetic Centre, with some smaller specialized laboratories carrying out some genetic research at Kuwait University; other private genetic laboratories exist for specific clinical tests but are also limited.

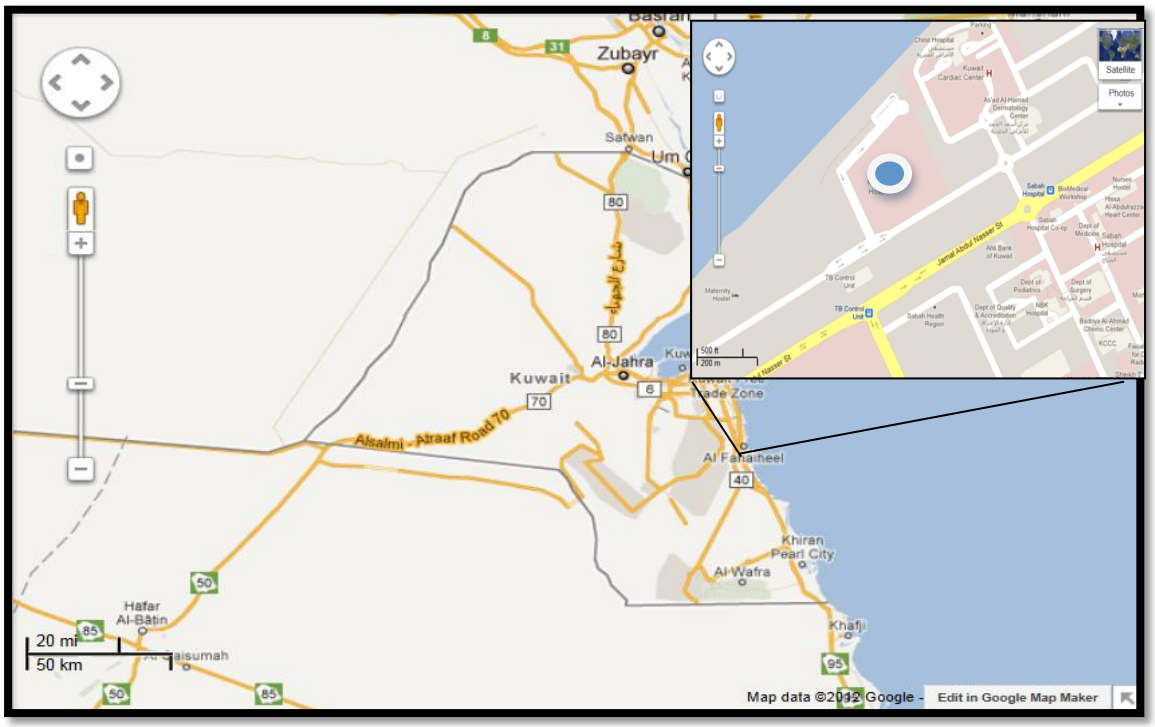


Figure 8: Location of Kuwait Medical Genetic Centre (blue circle on the map) within Kuwait. Image constructed from google maps

2.3.2 Genetic services

The main service offered at the genetic centre, often the first point of testing is chromosomal analysis from peripheral blood lymphocytes of Giemsa-banded (G-banding) chromosomes. Two banded karyotyping is performed by examining ten cells from each bringing a total of 20 counted cells with a turnaround time of 2-7 days.

Other staining methods for karyotyping can be applied such as C-banding, Q-banding, NOR and DAPI but are not routinely used and only performed if necessary. If no chromosomal abnormality is detected the case is then referred to analysis through specific FISH probes or to further molecular testing depending on the suspected disorder.

2.4 Consanguinity in Kuwait

In the mainly Muslim country of Kuwait, cousin marriages are permitted and first cousin or even double cousin unions are common. Such is the case of many neighbouring countries such as Saudi Arabia. Cousin marriages are often preferred to maintain wealth within the family and the purity of the breed within certain traditional Bedouin tribes. Like the Saudi counterparts high rates of consanguinity creates genetic isolates in which a founder recessive mutation can rise to a high frequency. In Saudi Arabia a first cousin marriage rate was reported at 57.7% and reaching more than 80% in certain areas (Al-Owain, Al-Zaidan et al. 2012). Reported consanguinity in Kuwait ranged from 44.8% to 54.3% across three different studies (Al-Kandari and Crews 2011)

Consanguinity increases the coefficient of inbreeding, which is defined as the percentage of the genome that is identical by descent or homozygous, this in turn increases the likelihood of a pathogenic mutation being found in a homoallelic state and a familial disorder is most likely to be caused by an autosomal recessive mode of inheritance (Alkuraya 2010). This makes consanguineous families useful in homozygosity mapping to identify causative mutations; this will be discussed further chapter 6.

A study by Al-Kandari and Crews conducted on 9104 married Kuwaiti females reported a significant difference in the incidence of physical and intellectual disability between the offspring of first-cousin marriages compared to non-consanguineous marriages, while no significance difference was observed between 2^{nd} or 3^{rd} cousin marriages compared to non- consanguineous counterparts (Al-Kandari and Crews 2011).

There are several examples of autosomal recessive disorders that have been identified in Saudi and Kuwaiti patients, such as the homozygous 12bp deletion in exon 3 of the TBCE gene leading to Sanjad-Sakati syndrome (Al-Owain, Al-Zaidan et al. 2012).

Autozygosity mapping was applied to a large consanguineous Pakistan family with non-specific ID, which lead to the identification of the 11p15-tel locus as a causative region (Rehman, Baig et al. 2011), while another study on 64 Syrian consanguineous families with non-specific ID revealed 11 new loci (Abou Jamra, Wohlfart et al. 2011).

The focus of this thesis is on the selection of consanguineous ID families with the suspected underlying mode of inheritance being autosomal recessive.

2.5 Family Selection

The reasons for families to get their children tested for disabilities include admission into specialist schools and claiming disability benefits. Genetic testing is requested for elder children and sibs to identify a genetic factor either to assist in a planned pregnancy or for pre-marital counselling.

The samples used in this study can be divided into two sets, one for non-specific intellectual disability (ID), which were selected based on the criteria below, and the other for families with one member with Rubinstein Taybi Syndrome (RTS), these were selected by the clinician's pre-diagnosis of the disorder based on distinct clinical features (See **Table 2** for list of samples). For most ID and some RTS blood samples collected, DNA was extracted in the laboratory by the author. For samples in the DNA bank, consent was taken before including them in the study. The author was involved in the selection process and even contacted some of the selected families from the registry books, searching patient files and attended most of the genetic counselling sessions and asked for consent to include members from the selected families in the study.

2.5.1 Selection criteria for non-specific ID

Families were selected from the Kuwait Medical Genetic Centre Patient registry by the author under the supervision of the Clinical Geneticist (Dr Makia Marafie) for inclusion in this study based on the following criteria:

- Consanguineous parents (either first cousins or distant relatives)
- Parents with no reported intellectual disability
- Have at least one affected sib with non-specific intellectual disability
- Affected individuals have normal karyotype
- Tested negative for FRAX-A & E mutation expansion.
- IQ score less than 70 using the Wechsler Intelligence Scale for Children (WISC)
- Parents consenting to samples being used in research

2.5.1.1 Intelligence score

The primary instrument used to assess IQ and diagnose ID is the **Wechsler Intelligence Scale for Children** (**WISC**) (Wechsler 2004). This is used for children between the ages of 6 and 16 inclusive and can be completed without reading or writing. The WISC takes 65–80 minutes to administer and generates an IQ score, which represents a child's general cognitive ability. Standard categories for classification of ID are shown in **Table 1**.

Table 1 Classification of ID by IQ levels.

Class	**IQ**
Profound intellectual disability	**Below 20**
Severe intellectual disability	**20-34**
Moderate intellectual disability	**35-49**
Mild intellectual disability	**50-69**
Borderline intellectual functioning	**70-84**

2.5.2 Family pedigrees

Detailed descriptions are given below for the 16 pedigrees selected for ID and the 10 selected RTS pedigrees. Clinical information was extracted from medical notes for each pedigree. DNA samples were available from individuals that are labeled in the pedigrees. Each individual was given a unique number, which is used in labelling the pedigree for easy reference. All pedigrees were created using Progeny 7 software. For marital relationships a double line is used to refer to consanguinity through first, or second cousin marriage or being related few generations back. Intelligence score for individuals was determined from the Wechsler Intelligence Scale for Children (WISC) and was performed by psychologists at the Genetic Centre.

Pedigrees show individuals affected with ID (MR=Y), with the proband indicated by an arrowhead as exemplified in **Figure 9.** below, females are referred to with a circle icon and males with a square icon. A double line is used to indicate consanguinity (seen in other pedigrees)

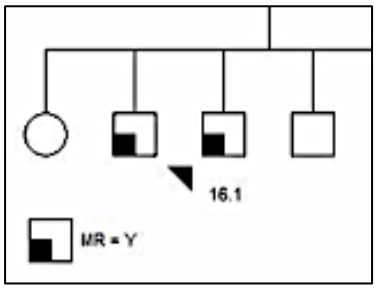


Figure 9: showing a section of a pedigree with a key

The level of ID on the samples available is listed in table 2 based on pedigrees and the information available. A similar list is generated on table 3 for RST samples.

Table 2 ID level of all ID samples collected and their sibs when known, number of affected sibs, family size, and degree of consanguinity.

Pedigree	Number of affected sibs	Individual labeled on the family pedigree with available DNA sample	Level of ID (inclusive of individuals not labelled on the pedigree figures)	Total number of sibs (affected and unaffected)	Degree of consanguinity
1	4	1.1 1.2 - -	Mild Mild Borderline Unknown	6	Distant cousins
2	2	2.1 2.2	Profound Moderate	6	Distant cousins
3	2	3.1 3.2	Severe Moderate	12	Distant cousins
4	3	4.1 - -	Borderline Moderate Severe	6	1st cousins
5	2	5.1 -	Mild Borderline	7	1st cousins
6	4	6.1 6.2 - -	Mild Mild Unknown Unknown	9	1st cousins
7	3	7.1 7.2 7.3 7.4 7.5	Severe Unaffected Moderate Moderate Unaffected	8	Distant cousins
8	3	8.1 8.2	Moderate Mild	6	Distant cousins
9	2	9.1 9.2 9.3 9.4 9.5	Unaffected Mild Unaffected Severe Unaffected	7	1st cousins
10	2	10.1 10.2 10.3 10.4 10.5	Moderate Mild Unaffected Unaffected Unaffected	3	1st cousins
11	3	11.1 11.2 11.3	Severe Severe Unknown	12	Distant cousins
12	2	12.1 -	Severe Unknown	6	2nd cousins
13	2	13.1 13.2	Severe Mild	7	1st cousin
14	4	14.1 14.2 14.3 -	Severe Severe Severe Severe	9	Distant cousins
15	3	15.1 15.2 15.3	Severe Moderate Severe	6	Distant cousins
16	2	16.1 -	Moderate Moderate	6	Distant cousins

Table 3 ID level of all RTS samples collected and their sibs when known, number of affected sibs, family size, and degree of consanguinity.

Pedigree	Number of affected sibs	Individual labeled on the family pedigree with available DNA sample	Level of ID (inclusive of individuals not labelled on the pedigree figures)	Total number of sibs (affected and unaffected)	Degree of consanguinity
A	2	A.1 A.2	Unknown Unknown	6	Distant cousins
B	2	B.1 B.2	Moderate Mild	2	1st cousins
C	1	C.1	Unknown	1	Non consanguineous
D	1	D.1	Unknown	1	Unknown
E	1	E.1	Moderate	5	Non consanguineous
F	1	F.1	Mild	5	Non consanguineous
G	1	G.1	Mild	7	Non consanguineous
H	1	H.1	Severe	4	Non consanguineous
I	1	I.1	Unknown	4	1st cousins
J	2	J.1 J.2	Unknown Unknown	8	Distant cousins

2.5.2.1 Pedigrees of families with non-specific ID

2.5.2.1.1 Family 1

Family 1

Family 1 (**Figure 10**) is a Kuwaiti family with four affected sibs with non-specific familial intellectual disability. The eldest is the proband who passed away from leukemia. He had mild ID.

The second affected individual has borderline ID with an IQ of 70. He has urinary incontinence at night and requires assistance with daily activities. He is interactive and responsive when asked questions.

Individual 1.2 has mild ID with an IQ of 52. He was referred to the Genetic centre as having developmental delay and difficulties at age seven and having three other affected siblings.

Results for FRAX-A and E mutation expansion for all siblings were negative. In addition FISH results and chromosomal analysis using peripheral blood for standard G-banding revealed normal 46, XY male karyotype for all of the male sibs. As apparent from the pedigree below, there are further extended family members who are affected from ID but were not contacted due to matters of confidentiality. The parents of the proband are related but are not first cousins.

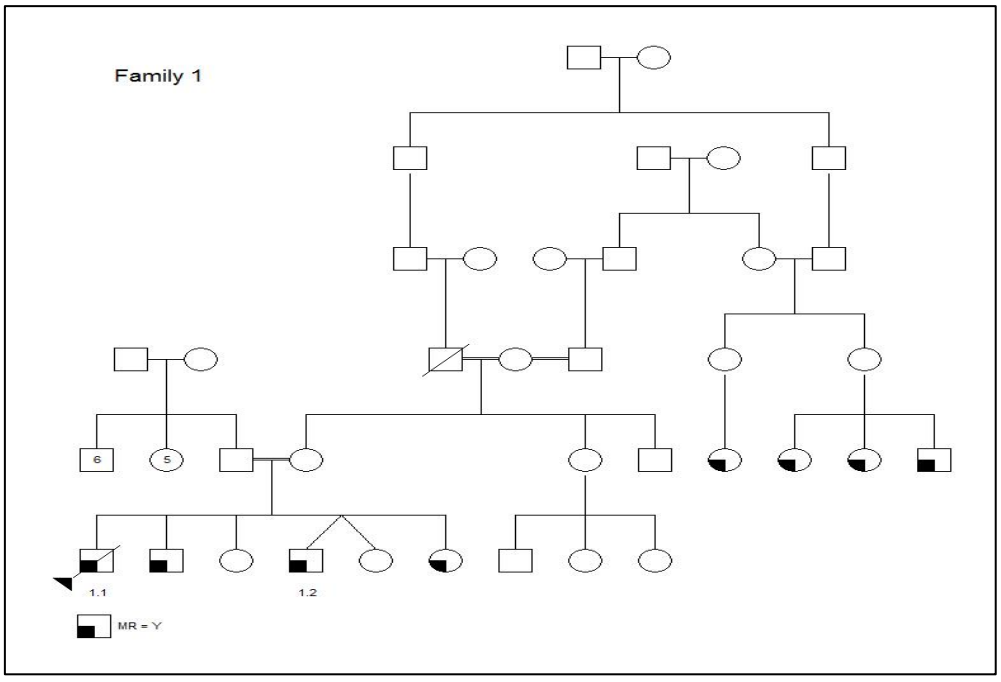


Figure 10: showing pedigree of family 1

2.5.2.1.2 Family 2

Family 2 (**Figure 11**) is of Bedouin (Mutairi) origin, consisting of two affected sibs and 4 unaffected sibs with parents who are second cousins. The affected sibs were ascertained in the Genetic Centre. The elder sister, individual 2.2, has moderate ID with some dysmorphic features that include a long face, microcephaly, delayed eruption of teeth, right Simian crease, bilateral clinodactyly of the fifth fingers and an abnormal gait. Her karyotype results, via G-banding, were of a normal female 46, XX. Amino acid chromatography result was normal as well. A visiting consultant to the centre suggested the possibility of Rhett's syndrome as a close diagnosis. Fragile X results for FRAX-A and E expansion was negative.

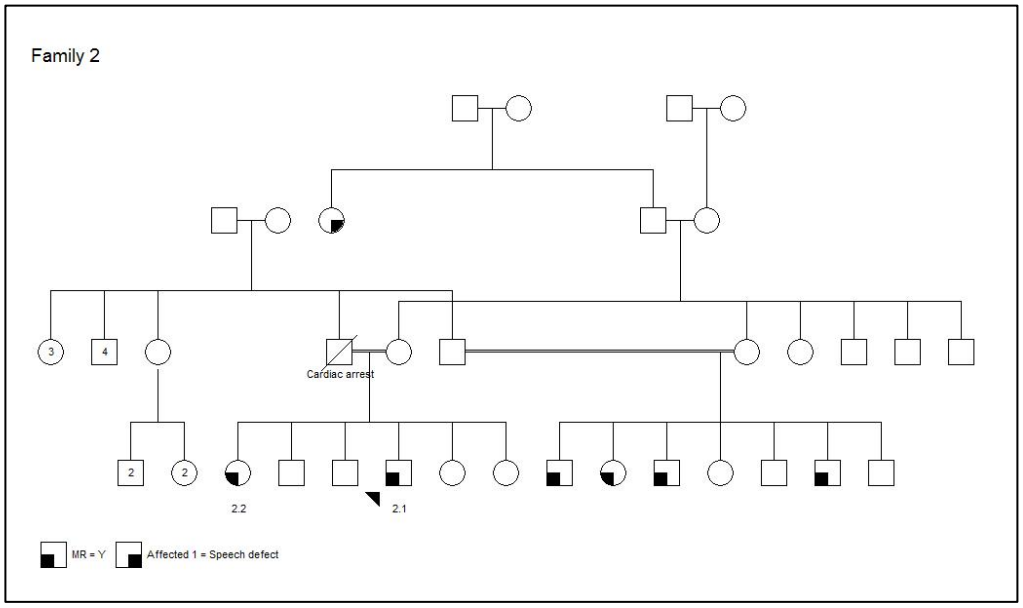


Figure 11: showing pedigree of family 2

The younger brother, individual 2.1, has severe ID. His birth weight was 2.9kg and was first reported, at age 7 months, by the Paediatrician of having psychomotor retardation with unsteady small head, spasm, with limited abduction of the legs and brisk lower limb reflexes. He has no dysmorphic feature and has a normal karyotype and amino acid chromatography results. Fragile X results for FRAX-A and E expansion was

negative. At age 8, he spoke few words and was wearing glasses. He needs assistance in personal hygiene. His ID level was changed in 2010 to profound ID.

2.5.2.1.3 Family 3

Family 3 (**Figure 12**) is a Kuwaiti family from the Bedouin Thafeeri group and consists of two affected sibs from a total of 12 sibs. Parents are consanguineous but are not closely related with a common ancestor being five generations back. However, two of their first cousins also have ID. Individual 3.1 has severe ID with an IQ of 34. Individual 3.2 has moderate ID, with an overall IQ of 40. Delayed speech was noted for both sibs. Results for FRAX-A and E mutation expansion for both siblings were negative. In addition, FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal 46, XY male karyotype for both sibs.

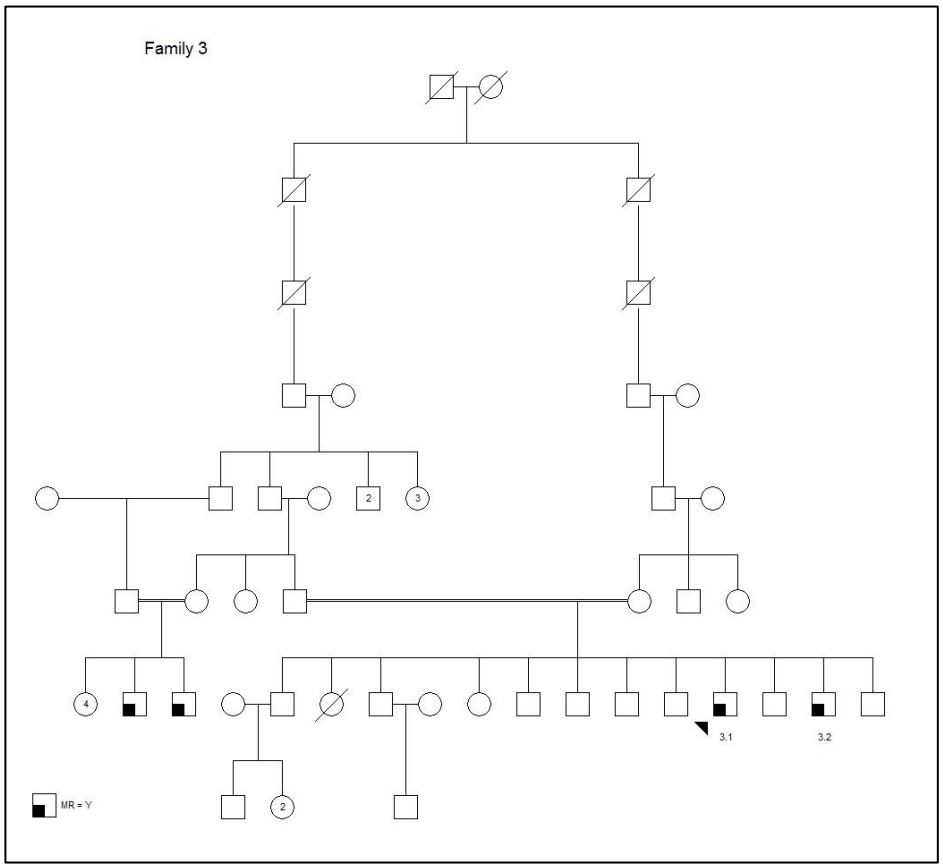


Figure 12: showing pedigree of family 3

2.5.2.1.4 Family 4

Family 4 (**Figure 13**) consists of three affected male sibs and three unaffected female sibs born to 1st cousin parents. The first affected sib has moderate ID with an overall IQ of 38. The second affected sib has severe ID with an IQ less than 35. While the third affected sib, individual 4.1 in pedigree 2.4, has borderline ID.

Results for FRAX-A and E mutation expansion for all siblings were negative. In addition, FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal 46, XY male karyotype for all sibs.

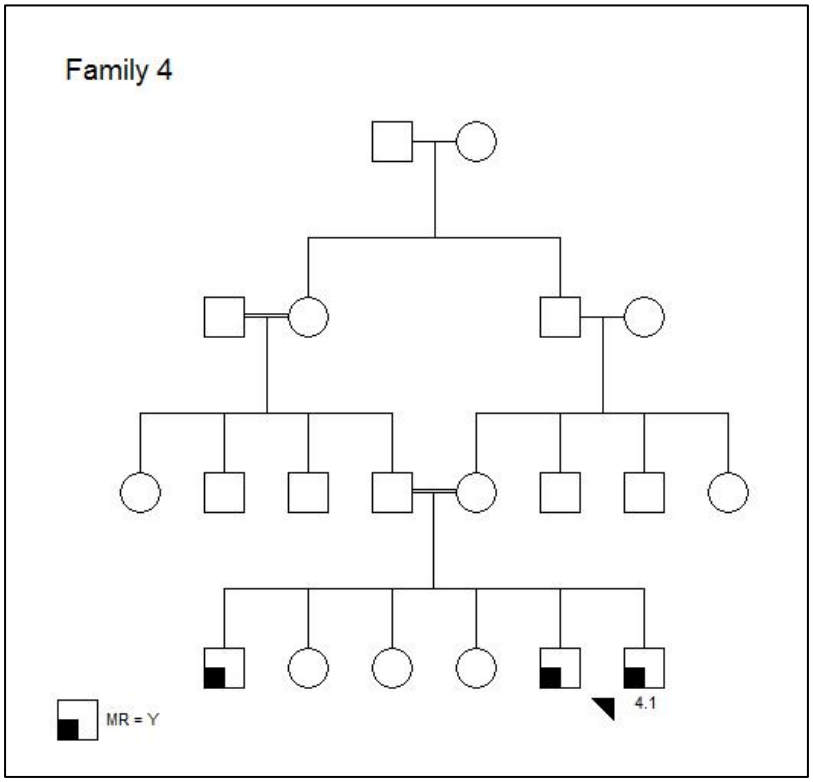


Figure 13: showing pedigree of family 4

2.5.2.1.5 Family 5

Family 5 (**Figure 14**) consists of two affected sibs with ID, and five unaffected sibs to consanguineous first cousin parents. The grandparents from the maternal side are also first cousins. Other disorders in the family include visual impairment, speech defect and constrictive paralysis.

Individual 5.1 has mild ID with an overall IQ of 65. This individual was referred and found to be highly myopic and was given spectacle correction by the optometrist. Her older sister has been classified as slow learner and later assessed and found to have an overall IQ of 74 indicative of borderline ID. The brother has speech difficulty and was classified as slow learner with an overall IQ of 73. In addition, two paternal uncles are reported to have ID.

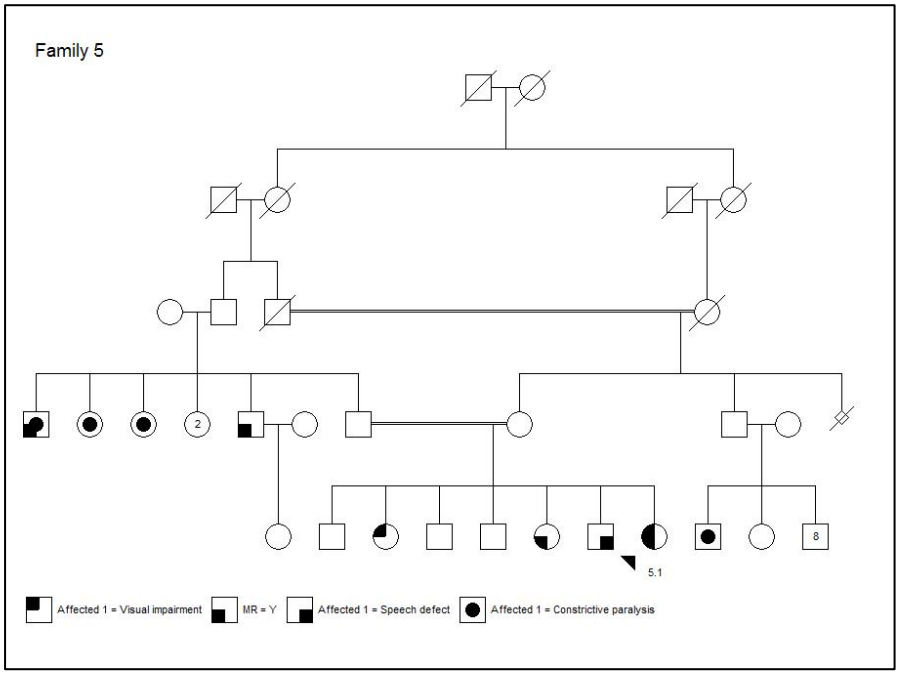


Figure 14: showing pedigree of family 5

Results for FRAX-A and E mutation expansion for all siblings were negative. In addition, FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype for all sibs.

2.5.2.1.6 Family 6

Family 6 (**Figure 15**) consists of nine children to first cousin parents with 4 children affected by ID. The mother had six spontaneous abortions. The mother and father also have consanguineous parents. In addition, four males on the maternal side have ID, two of which also have two affected sons each.

Individual 6.1 has mild ID with an IQ score of 61, while individual 6.2 has mild ID with an IQ score of 56. Results for FRAX-A and E mutation expansion for all siblings were negative. In addition, FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype for all sibs and the father.

Parents were contacted for further blood samples but failed to show up. Also, the parents did not want other family members to know that they visited the Genetic Centre.

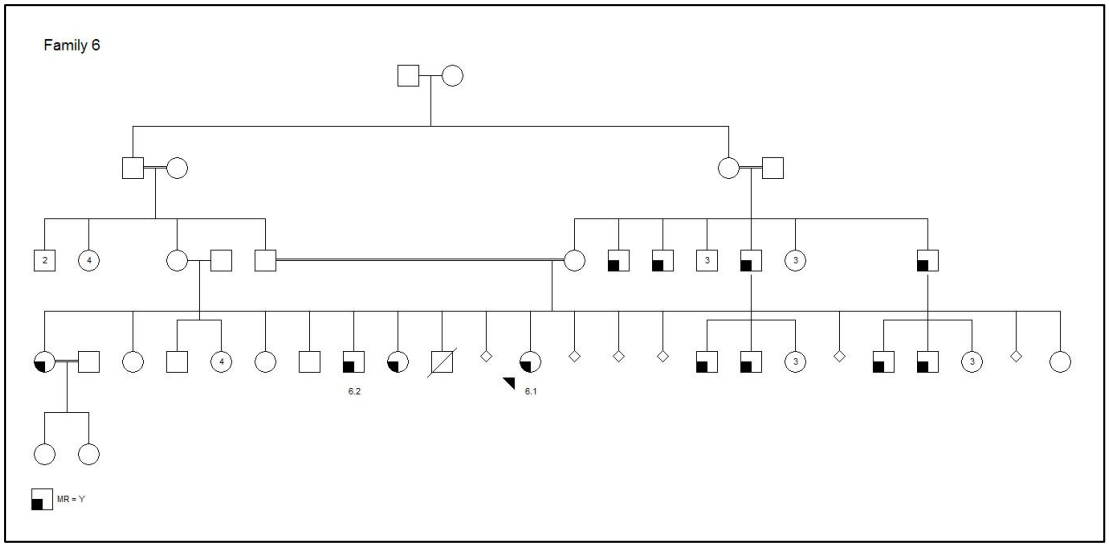


Figure 15: showing pedigree of family 6

2.5.2.1.7 Family 7

Family 7 (**Figure 16**) consists of eight offspring to consanguineous parents, 3 of which have ID. On the paternal side, two of the father's cousins are affected with ID.

Individual 7.4 has moderate ID, while individual 7.1 have severe ID, and individual 7.3 have moderate ID. Clinical records contained no information on IQ test results.

The results for FRAX-A and E mutation expansion for all siblings were negative. In addition, FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype for all sibs and the father.

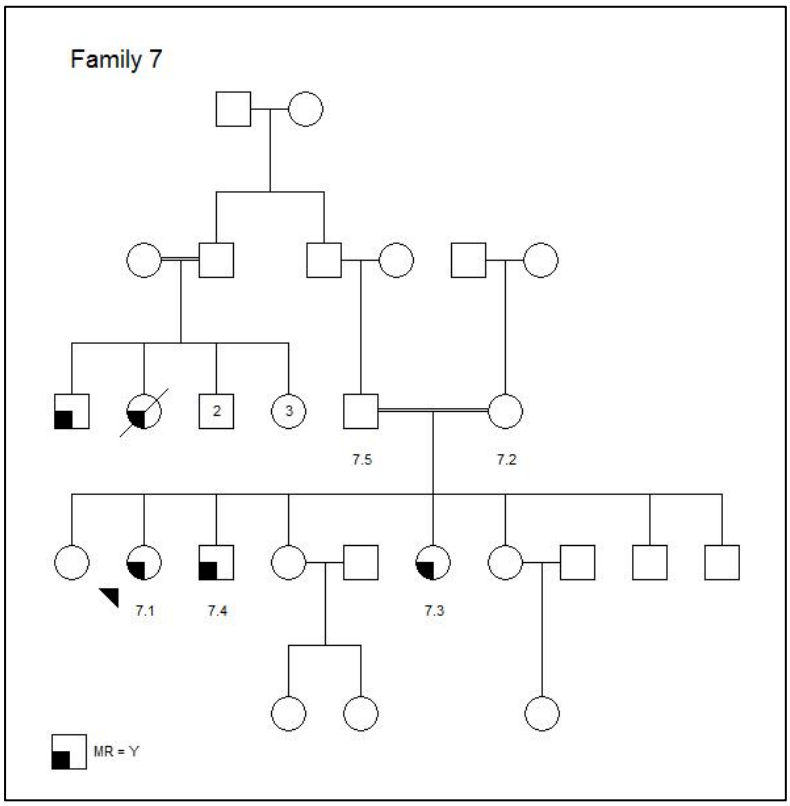


Figure 16: showing pedigree of family 7

2.5.2.1.8 Family 8

Family 8 (**Figure 17**) consists of six siblings to related parents. Three of the siblings have ID and one of the other siblings is deaf. Individual 8.1 has moderate ID with an overall IQ score of 45. Individual 8.2 has mild ID with an overall IQ score of 60.

Results for FRAX-A and E mutation expansion for both siblings were negative. In addition, FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype for both.

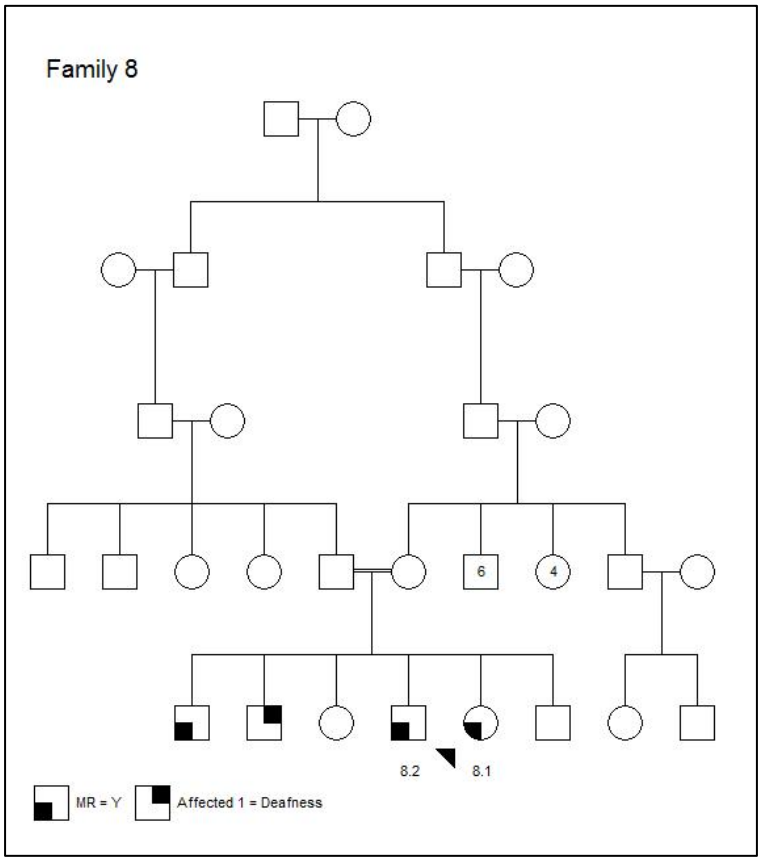


Figure 17: showing pedigree of family 8

2.5.2.1.9 Family 9

Family 9 (**Figure 18**) consists of seven sibs of a consanguineous marriage, with two individuals being affected by ID. Individual 9.4 has severe ID with an IQ score of 45, while his brother, individual 9.2, has mild ID. In addition, a cousin from the maternal side also has ID. Individual 9.4, a product of a full term pregnancy, was also reported to be hyperactive with speech delay, has familial hirsutism and with the only skeletal anomaly being a high arched foot (pes cavus). An MRI scan revealed prominent Cerebrospinal fluid **(**CSF) spaces, dilated ventricles and prominent sulci. In addition, he has a mild hearing loss in one ear.

Results for FRAX-A and E mutation expansion for both siblings were negative. In addition, chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype for both.

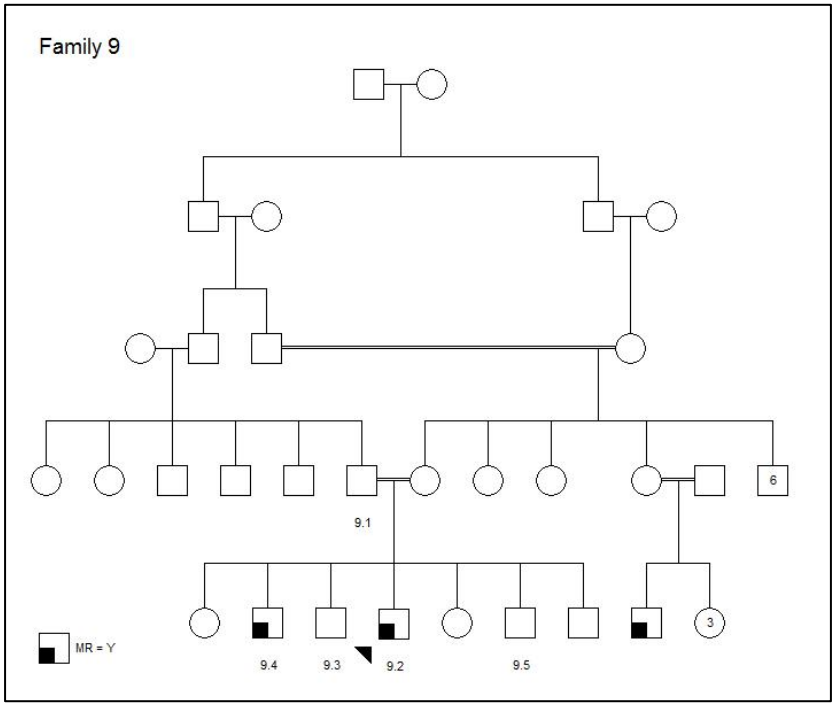


Figure 18: showing pedigree of family 9

2.5.2.1.10 Family 10

Family 10 (**Figure 19**) consists of 3 sibs to consanguineous parents. Two of the sibs are affected with ID as well as their paternal grandmother, six of their maternal cousins and three of their double second cousins. Individual 10.1 has moderate ID with an overall IQ score of 40, while his brother individual 10.2 has mild ID with an IQ score of 54. The second brother was born full term with birth weight of 2.7 kg. He started walking normally at age 1 year and a half and had speech delay.

Results for FRAX-A and E mutation expansion for both siblings were negative. In addition, chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype for both.

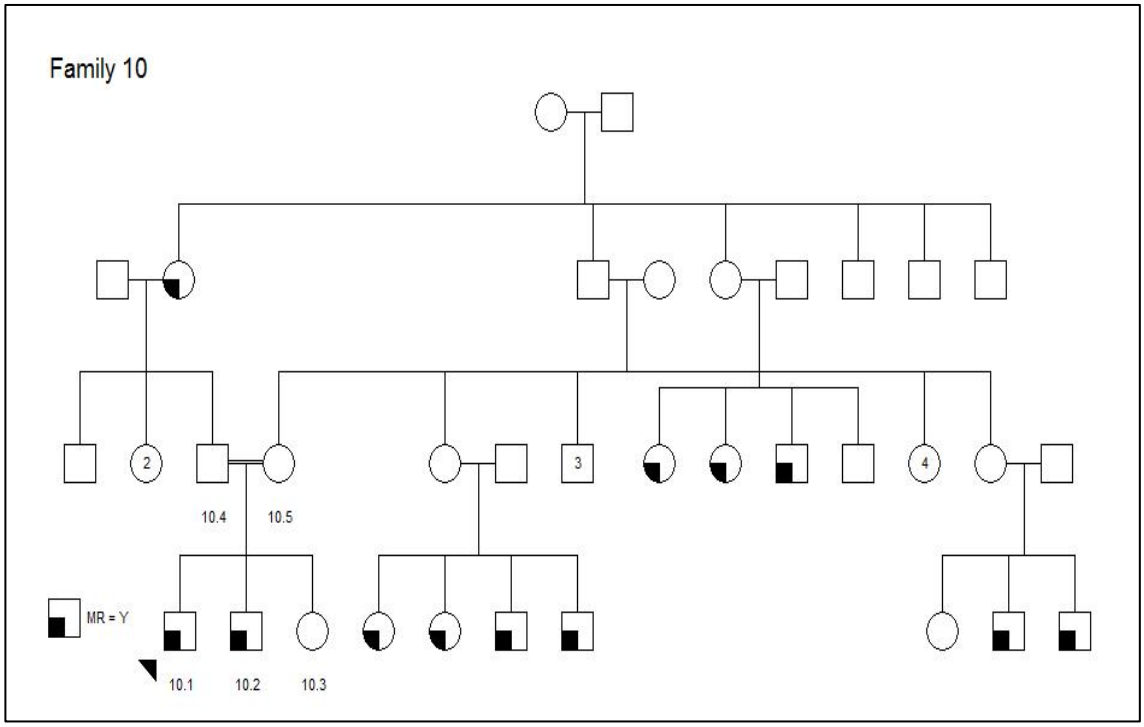


Figure 19: showing pedigree of family 10

2.5.2.1.11 Family 11

Family 11 (**Figure 20**) consists of 12 siblings from one marriage of a consanguineous couple, including three affected sibs and an additional affected sib from a previous marriage where the wife was also related. No other affected relatives were reported. Individual 11.1 has severe ID with an overall IQ score of 24. Individual 11.2 has severe ID as well. Severity was not reported for the other sib.

Chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype for individuals 11.1 and 11.2.

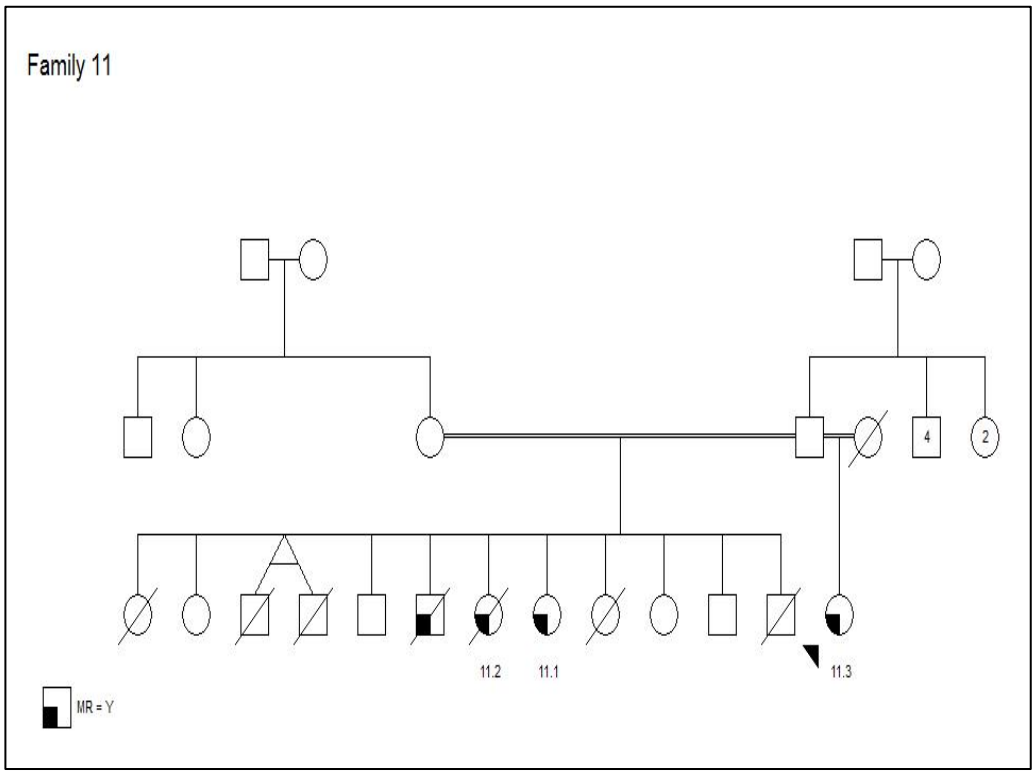


Figure 20: showing pedigree of family 11

2.5.2.1.12 Family 12

Family 12 (**Figure 21**) consists of six siblings to related parents. Two of the siblings have ID as well as one uncle and one cousin on the maternal side and five cousins from four different nuclear families on their paternal side. One cousin also has a speech defect in addition to ID. One cousin on the paternal side has psychosis while one maternal uncle and one of the siblings have sickle cell anemia. No other genetic disorders were reported. Individual 12.1 has severe ID with an IQ score of 44. He has long face and high arched palate. Chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype.

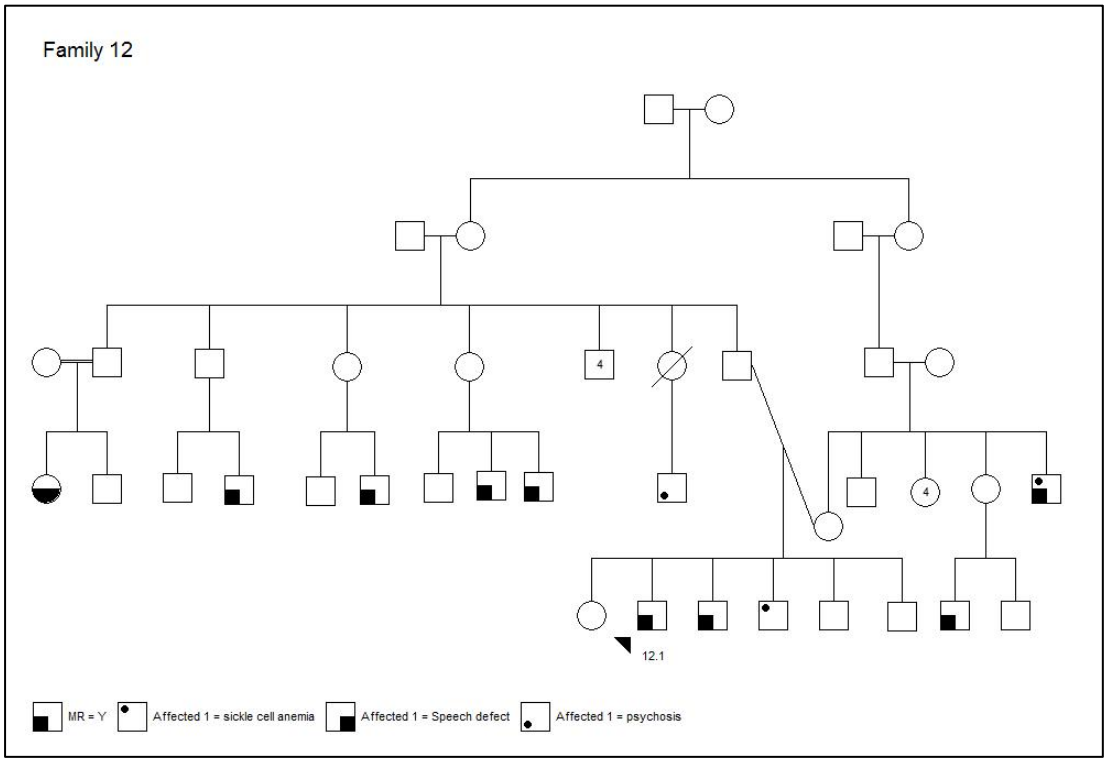


Figure 21: showing pedigree of family 12

2.5.2.1.13 Family 13

Family 13 (**Figure 22**) consists of seven children of a first cousin marriage with two affected female sibs. Both parents have consanguineous parents. The cousin of the two affected individuals is also affected with ID. Individual 13.1 has severe ID with an overall IQ of 31. Also, this individual has speech defect and speech therapy was recommended. In addition, individual 13.2 has mild ID with an IQ of 68.

Results for FRAX-A and E mutation expansion for both sisters were negative. In addition, FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal female karyotype 46,XX for both.

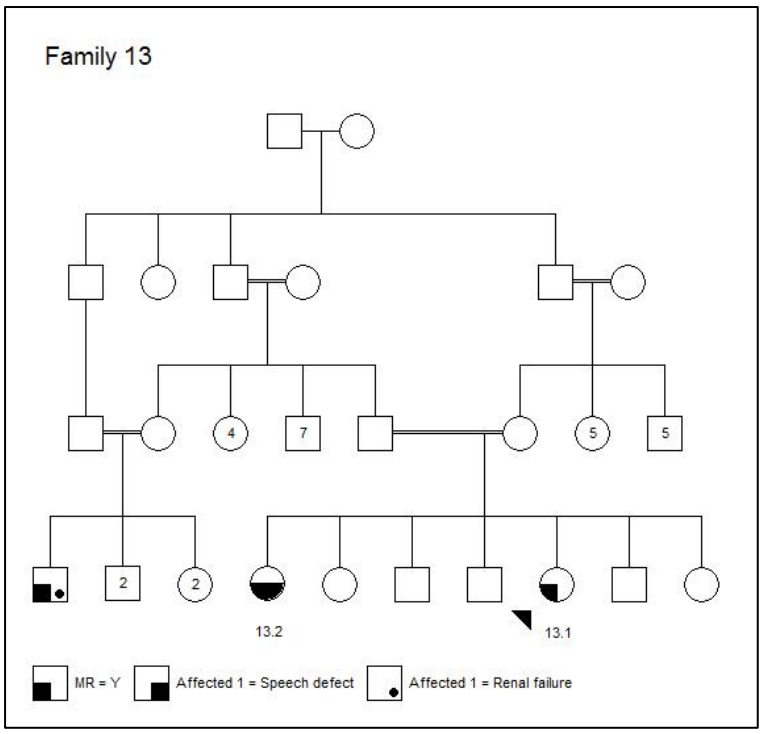


Figure 22: showing pedigree of family 13

2.5.2.1.14 Family 14

Family 14 (**Figure 23**) consists of 9 sibs to consanguineous parents, including four affected individuals. Two out of their five double cousins are also affected as well as their maternal aunt and their father's cousin. The mother had one spontaneous abortion. The sons, individual 14.1, individual 14.2 and individual 14.3 all have severe ID, with IQ scores less than 35. All of the affected children had developmental delays including speech delay.

Results for FRAX-A and E mutation expansion for all affected sibs were negative. In addition, FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype for all.

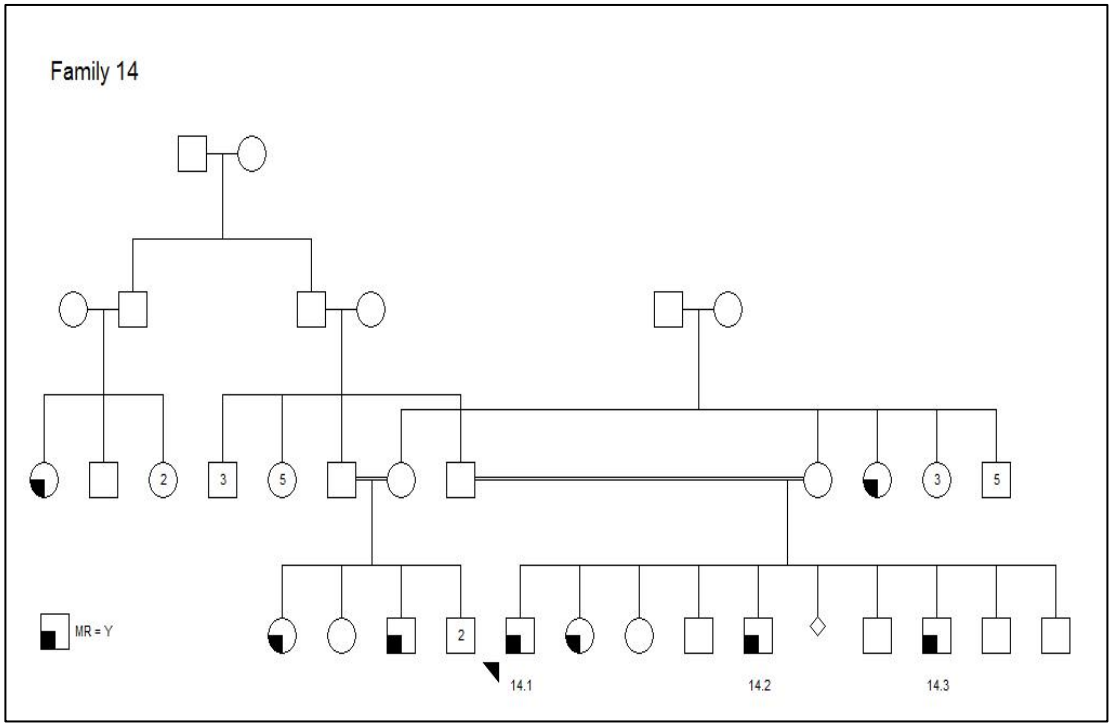


Figure 23: showing pedigree of family 14

2.5.2.1.15 Family 15

Family 15 (**Figure 24**) consists of 14 children to consanguineous parents, with six affected individuals with ID, two of which also have epilepsy. There is also one affected cousin. Individual 15.1 has severe ID with an overall IQ of 34, individual 15.2 has moderate ID with an overall IQ of 45, and individual 15.3 has severe ID.

Results for FRAX-A and E mutation expansion for all affected sibs were negative. In addition, FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype for all.

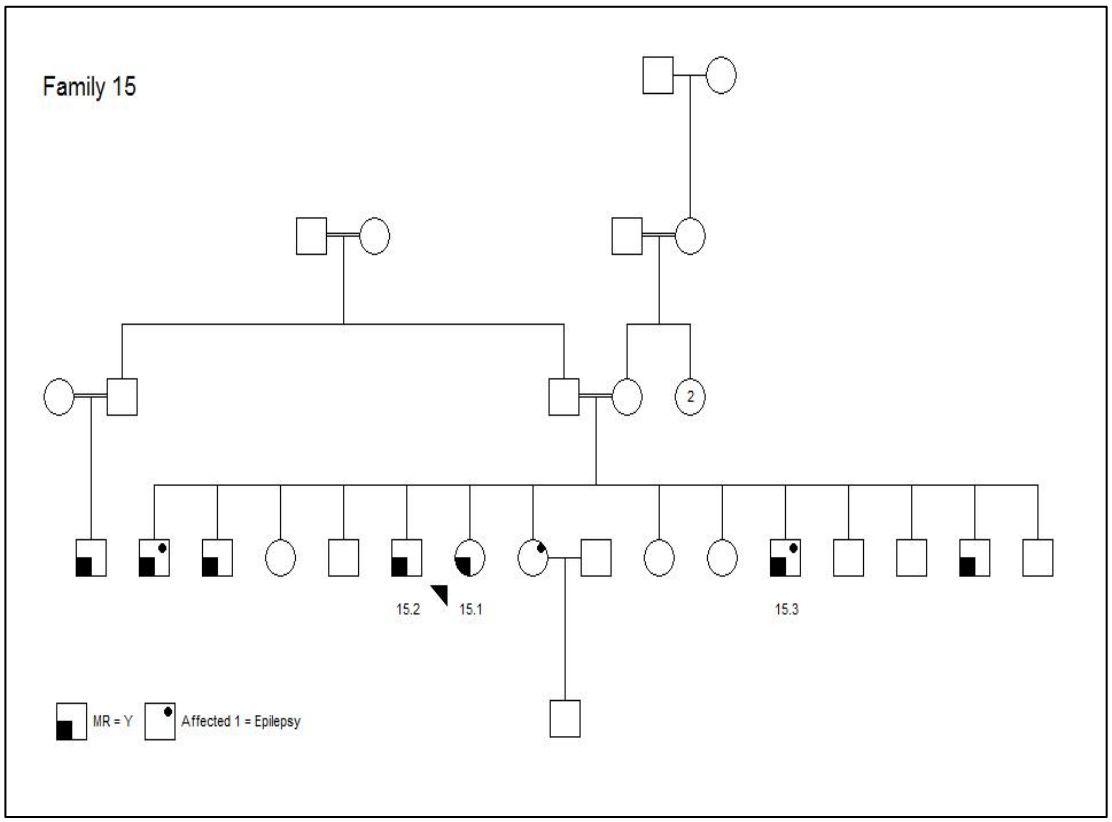


Figure 24: showing pedigree of family 15

2.5.2.1.16 Family 16

Family 16 (**Figure 25**) consists of six siblings, including two affected with ID, to consanguineous parents. Both affected males individual 16.1 and his brother have moderate ID with an overall IQ of 45.

Results for FRAX-A and E mutation expansion for both brothers were negative. In addition, FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype for both.

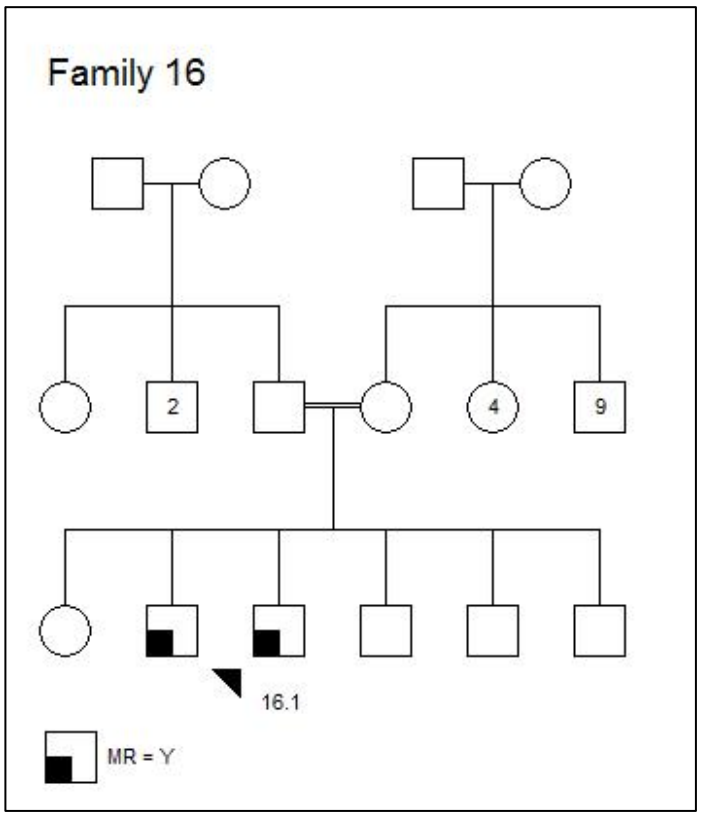


Figure 25: showing pedigree of family 16

2.5.2.2 Pedigrees of families with Rubinstein Taybi Syndrome (RTS)

RTS families selected based on clinical examination with distinct features as listed in the summary **Table 4**. The IQ score is included when available as intellectual disability is often found as part of the clinical features.

2.5.2.2.1 Family A

Family A (**Figure 26**) consists of six sibs to consanguineous parents of Bedouin origin and Saudi nationality, two of which, individual A.1 and A.2 have clinical features that include peaked nose, short stature, prominent ears, hyper mobility of joints, low hair line and ID. Upon medical examination multiple congenital anomaly/mental retardation (MCA/MR) syndrome was suggested with features being consistent with RTS. Chromosomal analysis using peripheral blood and using standard G-banding and also FISH technique revealed normal karyotype for both. Clinical features for both affected sibs are summarized in **Table 4**.

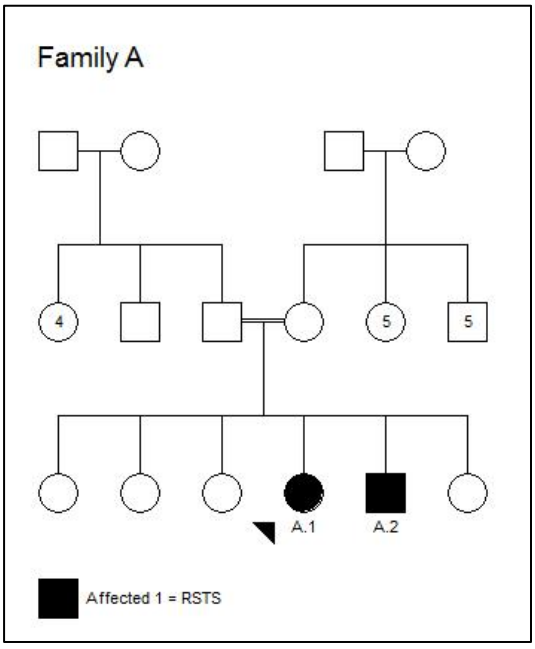


Figure 26 showing pedigree of family A

2.5.2.2.2 Family B

Family B (**Figure 27**) is of a consanguineous marriage. The clinicians at the Genetic Centre saw the proband, individual B.2, back in 2001. He was given an overall IQ of 63 that corresponds to mild ID. He was also diagnosed with hyperactivity, epilepsy, and hypothyroidism; he was on growth hormone therapy and was diagnosed with Rubinstein Taybi syndrome based on the clinical features in the summary **Table 4**. He was tested negative for FRAX-A and E. In addition, a CT brain scan showed normal results.

His sister, individual B.1, was also diagnosed with moderate ID based on an IQ of 48 and the clinical features are also summarized in **Table 4**. DNA samples from both children were sent to a laboratory in Germany for CREBBP gene (associated with RTS) sequencing and it was reported that no mutations were found in the CREBBP gene in those individuals.

The mother was trying to conceive again and was referred to IVF but required the confirmation of a causative gene mutation. She had three miscarriages.

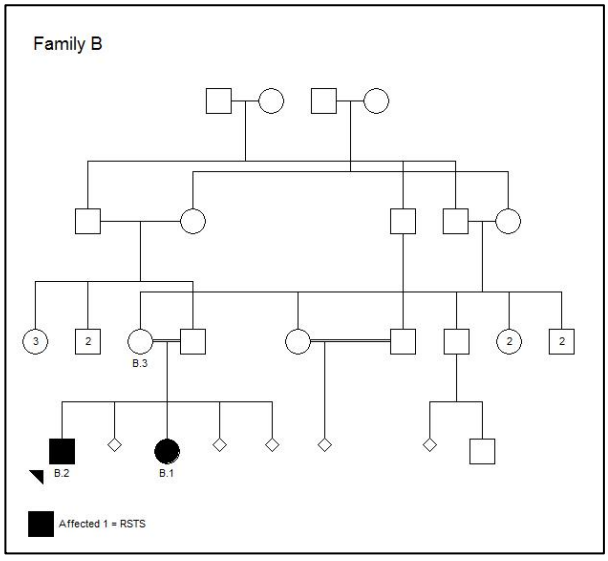


Figure 27 showing pedigree of family B

2.5.2.2.3 Family C

Family C (**Figure 28**) consists of a single child of unrelated parents. The proband was born of full-term normal vaginal delivery at a weight of 3kg. At age two her medical report included global developmental delay, microcephaly, subnormal growth, gastro-oesophageal reflux disorder (GERD) and the possibility of RTS. Chromosomal analysis using FISH revealed a normal karyotype with no 16p13.3 deletion. Clinical features are included in **Table 4**. In addition, it was noted that her broad thumb and toes resembled those of her father.

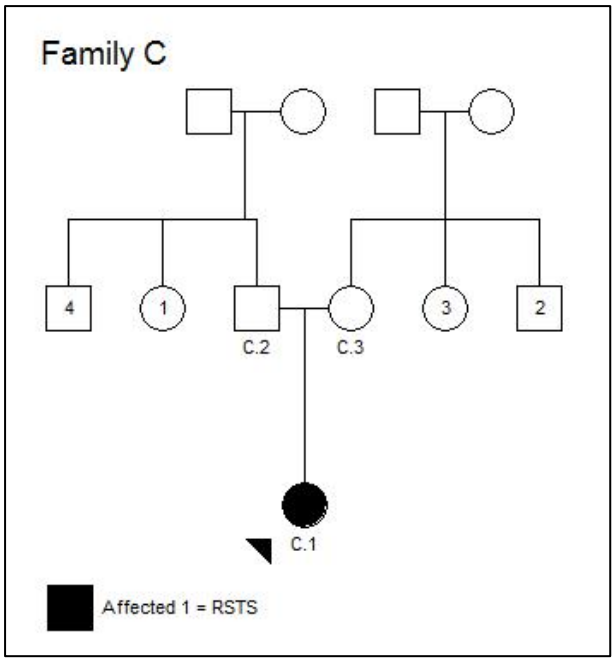


Figure 28 showing pedigree of family C

2.5.2.2.4 Family D

For family D (**Figure 29**) detailed information on the clinical features was missing from medical records except for broad thumbs and hallux.

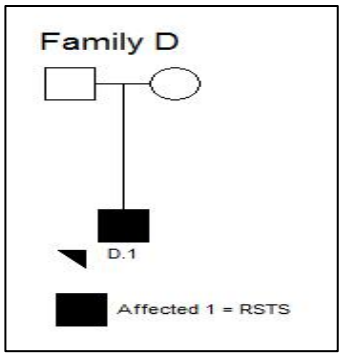


Figure 29 showing pedigree of family D

2.5.2.2.5 Family E

Family E (**Figure 30**) consists of five sibs to unrelated parents with one child, individual E.1 being affected with RTS and moderate ID. The proband was born after full term at 3.85 kg to a father at age 30 and mother 36 years of age. There were no postnatal complications. She was referred to the genetic centre for mild microcephaly and developmental delays. Chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype of 46,XX. Clinical features are summarized in **Table 4.**

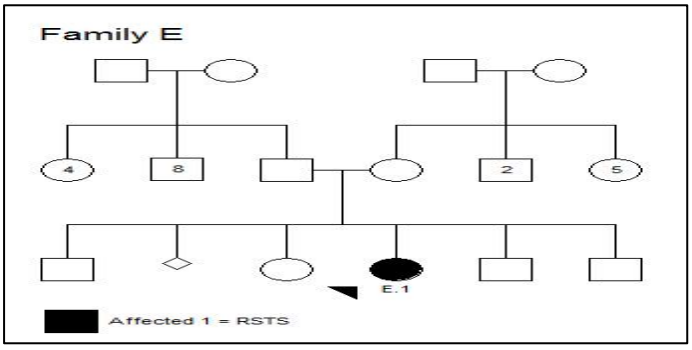


Figure 30 showing pedigree of family E

2.5.2.2.6 Family F

Family F (**Figure 31**) consists of five sibs to unrelated parents, the mother being Kuwaiti. The affected sib, individual F.1 has mild ID in addition to other RTS clinical features as listed in **Table 4.** His left kidney was removed at age 7 years.

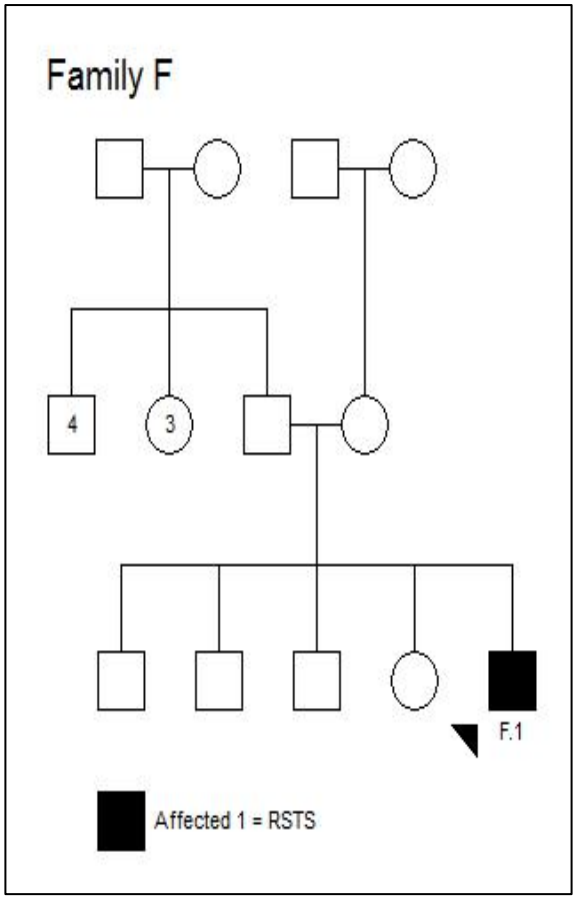


Figure 31 showing pedigree of family F

2.5.2.2.7 Family G

Family G (**Figure 32**) consists of seven sibs with the youngest, individual G.1, being affected with RTS, having dysmorphic features, congenital heart problem, mild ID, as well as severe visual impairment. His birth was normal after full term at weight of

3.45kg a 42-year-old father and 36-year-old mother. Chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype of 46,XY. Skeletal survey and audiometry was normal. IQ was reported at a score of 58. Clinical features are detailed in **Table 4.** Another suspected syndrome is CHARGE syndrome (The letters stand for: coloboma of the eye, heart defects, atresia of the nasal choanae, retardation of growth and/or development, genital and/or urinary abnormalities, and ear abnormalities and deafness).

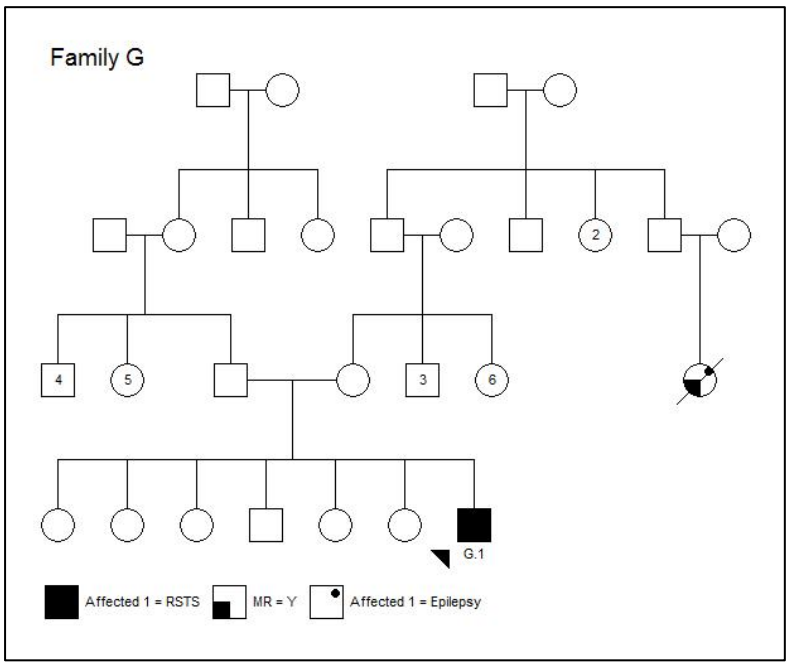


Figure 32 showing pedigree of family G

2.5.2.2.8 Family H

Family H (**Figure 33**), the proband is a triplet with severe mental and physical disability in addition to blindness and deafness. Other clinical features include severe hypotonia and dysmorphic features including broad thumbs and toes. The individual was a preterm at 32 weeks with birth weight of 1.07Kg, admitted to NICU for 3 months. Sequencing of the CREBBP gene for all exon 1 to 31 was carried out in Germany and no

mutation was reported. Chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype of 46,XY.

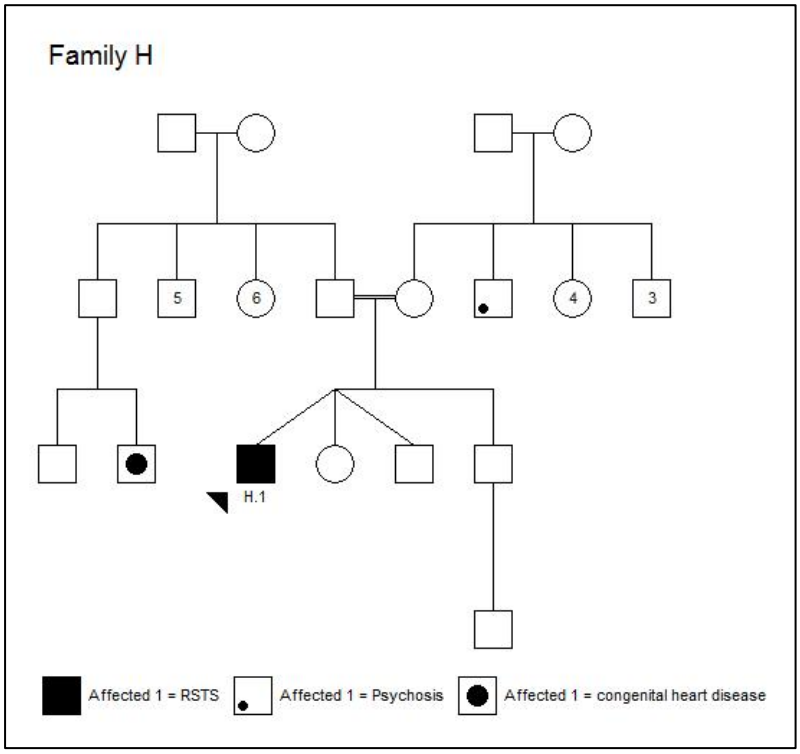


Figure 33 showing pedigree of family H

2.5.2.2.9 Family I

Family I (**Figure 34**) consists of four sibs to consanguineous parents including one affected child, individual I.1, with RTS. The proband was diagnosed at age 6 months to have hypotonia. The child completed full term and there were no postnatal complications. Clinical features include broad toes and thumbs.

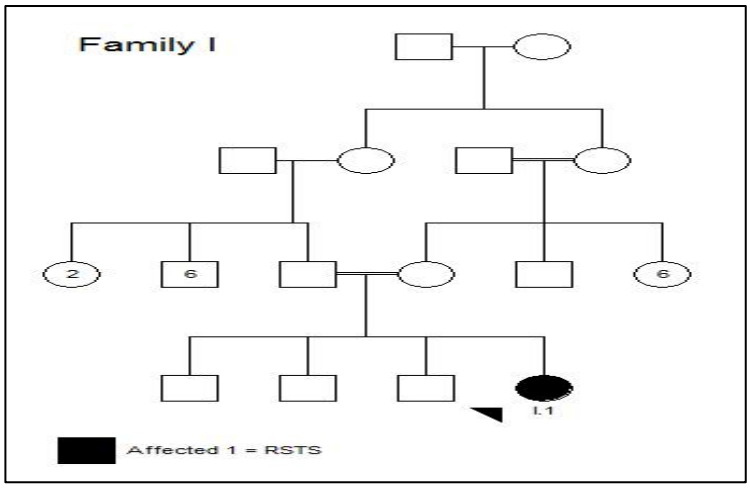


Figure 34 showing pedigree of family I

2.5.2.2.10 Family J

Family J (**Figure 35**) consists of eight siblings to consanguineous parents. Two of the sibs have RTS. The FISH results and chromosomal analysis using peripheral blood and using standard G-banding revealed normal karyotype with no 16p13 deletion (which encompasses the CREBBP gene) in both sisters. Clinical features for both affected sibs are in the summary **Table 4.**

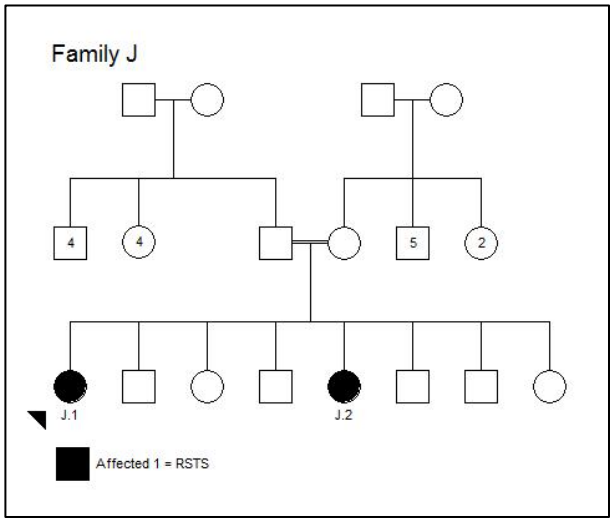


Figure 35 showing pedigree of family J

Table 4 showing a summary of the clinical features for each affected RTS individual

Clinical feature	A.1	A.2	B.1	B.2	C.1	D.1	E.1	F.1	G.1	H.1	I.1	J.1	J.2
Short stature, proportionate	■	■					■		■			■	■
Generalized hirsutism	■	■							■			■	■
Frontal upsweep/cowlick		■					■						
Low frontal hairline	■	■					■		■			■	■
Dysplastic ears							■		■				
Low-set ears							■	■					
Posteriorly rotated ears				■			■	■					
Deep-set eyes	■		■					■	■			■	■
Hypertelorism	■		■	■	■		■	■	■			■	■
Coloboma of iris									■				
Coloboma of retina/choroid									■				
Myopia		■		■									
Strabismus	■	■	■	■			■				■		
Arched eyebrows	■											■	■
Thick eyebrows	■											■	■
Long/prominent eyelashes	■	■			■							■	■
Ptosis of eyelids	■	■					■			■			
Palpebral tissues slant down	■	■	■	■									
Epicanthic folds					■		■			■			
Large nose	■	■	■	■	■		■	■	■				
Convex/beaked profile of nose	■	■					■		■	■		■	■
Columella below alae nasi	■	■		■			■	■					
Deviated nasal septum	■	■	■				■	■	■				
Small mandible/micrognathia	■	■		■			■					■	■
Hypoplastic maxilla -excluding molar region	■	■					■	■					
Microstomia		■											

Clinical feature	A.1	A.2	B.1	B.2	C.1	D.1	E.1	F.1	G.1	H.1	I.1	J.1	J.2
Thin upper lip	■	■		■			■			■		■	■
Narrow palate	■	■	■	■			■	■					
Cleft uvula							■						
Abnormally shaped teeth	■	■	■	■			■	■				■	■
Low posterior/trident hairline	■	■	■	■			■	■				■	■
Scoliosis							■						
Atrial septum defect								■	■				
Congenital cardiac anomaly, unspecified			■					■					
Patent ductus arteriosus				■				■					
Ventricular septal defect									■				
Vascular ring								■					
Hypospadias									■				
Cryptorchid testes				■					■				
Renal agenesis								■					
Skin syndactyly of fingers	■		■				■	■	■				
Wide finger tips	■	■	■	■			■						
Wide phalanges	■		■	■						■			
Broad thumbs	■	■	■	■	■	■	■	■	■		■		
Perthe's/dysplastic hip		■											
Broad hallux	■		■	■	■	■	■	■	■	■	■	■	■
Polydactyly/bifid hallux		■											
Absent/hypoplastic thymus	■												
Mental retardation/developmental delay	■	■		■			■	■	■	■		■	■
Microcephaly	■	■					■		■			■	■
Obese		■											

2.6 Discussion

From a total of 16 ID and 10 RTS families selected, 53 samples were collected including unaffected individuals when possible. It was difficult to gather more DNA as patients did not come back even when an appointment was booked for blood extraction. This could be partly because the study did not have an immediate impact on the family. More families were selected initially but some declined as they felt that it was not worth going through the hassle of blood extraction nor did they want their children to miss school. ID level ranged from profound to severe, with variance found amongst sibs. Consent was given for all the samples used even if the blood was not initially extracted for the purpose of the study.

For RTS cases, diagnosis was based on clinical features only and some samples were sent for CREBBP sequencing abroad with no mutations reported. Although RTS is often considered as a de novo dominant disorder, three of the ten families in this study had two affected sibs, suggesting another mode of inheritance or perhaps a parental mosaicsm, which might not have showed full penetrance in one of the parents.

These samples were used to evaluate experimental design, laboratory techniques, procedures and workflow for the detection of the genetic cause for the disorder in question. New laboratory techniques were proposed for use in the Genetic Centre and these will be the subjects of the next chapters.

Chapter 3- Materials & Methods

Overview

This chapter describes the procedures used for DNA extraction using two methods, a) Phenol/Chloroform and b) Maxwell ® 16 instrument by Promega. Extracted DNA was then used for one or several of the other methods, which include Cytogenetic arrays using Affymetrix 2.7M cytogenetic arrays, Direct sequencing using the ABI 3130 Genetic Analyzer (mainly for Rubenstein Taybi cases), and Multiplex Ligation-dependant Probe Amplification kits supplies by MRC-Holland and run on the Beckman Coulter CEQ^{TM} Fragment Analysis System. Three MLPA kits were used for duplication deletion screening and they are P245-A2 microdeletions, P036-E1 Telomere-3, P070-B2 telomere-5.

Results are listed in the individual chapters. All laboratory methods applied for the DNA samples were carried out by the author. Some blood samples were extracted by the author, except when samples were collected without the presence of the author, in which a laboratory technician would perform the extraction and label the samples ready for use later by the author. DNA samples from two individuals were sent for exome testing and analysis at Guys hospital, London.

3.1. DNA Extraction from whole blood

Blood was drawn from cases and relatives described in chapter two at Kuwait Medical Genetic (KMGC) by a nurse at a volume of 5ml of blood in a 4ml lavender-top vacutainer (K2E 7.2mg, Beckton-Dickinson, Plymouth; EDTA anticoagulant). DNA was then extracted from whole blood using one of two protocols listed below. The phenol/chloroform extraction protocol will be referred to as protocol 1, while Maxwell 16 DNA extraction will be referred to as protocol 2.

3.1.1 The phenol/chloroform method (protocol 1)

This is a two-day method based on phase separation by centrifugation, after mixing the lysed blood cells with phenol/chloroform, the mixture is separated into aqueous phase, interphase, and organic phase upon centrifugation. DNA would be in the aqueous phase and protein would be in the organic phase. This was the main method of choice for all extractions at the Kuwait Medical Genetic Centre due to its high DNA yield.

3.1.1.1 Solution preparation

Cell lysis buffer

Cell lysis buffer was prepared by adding the following;

NH_4Cl 8.29 g [final concentration 155 mM],

$KHCO_3$ 1 g [final concentration 10 mM]

Na_2EDTA 0.034 g or 200 µl EDTA 0.5 M [final concentration 0.1mM]

Then the volume was made up to 1000 ml (as marked by the volume label on the bottle) with distilled water and pH was adjusted to pH7.4 with 1M HCl or NaOH.

Ready made solutions:

TEN buffer (Promega, Leiden, Netherlands) one tablet in 100ml of distilled water

Protease K (30u/mg, Promega, Leiden, Netherlands)

Phenol/chloroform mix: also known as Phenol/chloroform/isoamyl Alcohol, 1 Phase 25:24:1 (Amresco, Ohio, USA)

3.1.1.2 Day 1: Cell, nucleus lysis and protein digestion

Whole blood was transferred to a 50ml tube and lysis buffer was added to the 50ml mark and mixed by tilting the tube up and down. When fresh blood was used, it was incubated at room temperature for 10 minutes before centrifugation.

The blood mixture was then centrifuged at 3000 rpm for 15 minutes. The supernatant was then discarded in a special container for hazardous liquids (for safe disposal of the blood sample) and 3ml of TEN buffer (a buffer solution to solubilise DNA and protect it from degradation) was added and mixed well, followed by 150 µl of protease K (30u/mg, Promega, Leiden, Netherlands), mixed, and finally 300 µl of 10% SDS (Sodium dodecyl sulfate, Promega, Leiden, Netherlands) was added and mixed before incubation at 37ºC overnight.

3.1.1.2 Day 2: DNA precipitation

After the mixture from day 1 had been incubated overnight, 1.5ml of saturated sodium chloride (6M NaCl) was added and mixed before centrifuging at 3000 rpm for 15 minutes. The supernatant was then transferred to a fresh 15ml tube and 1-1.5ml of phenol/chloroform mix (Amresco, Ohio, USA) was added and mixed gently for 2 minutes, then centrifuged at 3000 rpm for 15 minutes. The supernatant was then transferred to a new tube and two times the volume of cold ethanol was added and quickly mixed to precipitate DNA, which appeared as a white threaded lump.

The DNA was then removed using a hooked pasture pipette and washed by dipping the hook in 70% ethanol twice and then leaving the pellet to air dry for 2 minutes. DNA was then dissolved in a 1.5ml tube containing 300-500 µl of T.E. buffer (1X, 100ml, Promega, Leiden, Netherlands) and left at room temperature overnight and then stored at -20ºC for use when needed.

3.1.2 Using Maxwell® 16 instrument by Promega (protocol 2)

This method was introduced at the Genetic Centre as a robust and less hazardous method for DNA extraction.

The Maxwell® 16 DNA cartridges kit is supplied with purification plungers, elution tubes and elution buffer. Fresh blood was used or stored at room temperature for no more than 24hours. The procedure was carried out using 400 µl of whole blood following manufacturer's instructions for DNA extraction from whole Blood.

Each supplied cartridge comes with the reagents ready in each well. The cartridge to be used is placed into the holder with the ridged side facing the numbered side of the rack. The seal is then removed and a plunger is placed into well #7 of the cartridge and the blood sample is added to well #1, as illustrated in **Figure 36**.

An elution tube was labelled and 400 µl of elution buffer (to elute the purified DNA) was used as supplied with the DNA purification kit (Promega, Leiden, Netherlands). The machine was then switched on and when the screen indicated "SEV", the Run/Stop button was pressed. DNA option was selected and then Blood protocol was also selected and again the Run/Stop was pressed and the door was opened when prompted.

The cartridges were transferred from the preparation racks to the machine platform and the elution tubes were placed in front of the cartridge in the matching elution tube

slot. The Run/Stop button was pressed again to start the DNA purification process. When completed the eluted DNA was then transferred to a fresh 1.5 µl tube labelled with the sample number. The samples were stored at -20 ºC ready for use.

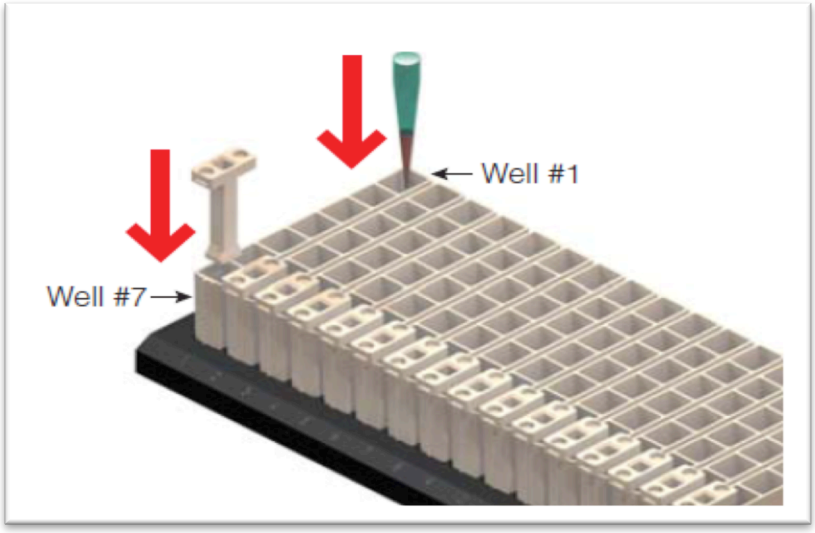


Figure 36 the cartridges and wells with well positions indicated, adapted from the Maxwell® 16 user manual

3.2 Sequencing of the CREBBP gene

Direct sequencing was carried out to identify single base mutations within the CREBBP gene, using the ABI 3130 Genetic Analyzer. The target DNA fragment is amplified and the product undergoes a second amplification process where the sequence is terminated by dideoxynucleotidesas producing different size fragments (sequence PCR) and the different DNA bases (adenine, guanine, cytosine, and thymine) are labeled with four different dyes. DNA Sequencer has a laser that emits at a wavelength absorbed by the florescent dye that is used to label the DNA, and an optical detector will detect the wavelength the dye fluoresces at. The different lengths of the fragments

produced are separated through a gel and the results are complied by the software to produce the DNA sequence.

Sequencing steps include, primer design and preparation, PCR reaction and product purification, Sequence PCR and fragments purification followed by direct sequencing with the genetic analyser.

3.2.1 PCR of CREBBP exons

3.2.1.1 Primer design

Primers for CREBBP gene were selected based on methods published by Coupry et al 2002. Primer sequences were checked with the PCR tool on the UCSC browser (www.genome.ucsc.edu). Using the In-Silico PCR human genome was selected and the forward and reverse primers sequences were inserted in the appropriate fields. The expected product for each primer pair was copied for future reference of the expected products (these are can be found in the appendix). The primer pairs selected are listed in ***Table 5***

Table 5 *below. Primer selection was based on exons with most reported mutations that cover the HAT domain.*

All primers were ordered with M13 sequence extensions as shown below with the forward sequences for all forward primers and the reverse sequence for all reverse primers.

M13-Fwd	5'-GTAAAACGACGGCCAGT-3'
M13-Rev	5'-AACAGCTATGACCATG-3'

Table 5 showing the list of forward and reverse primers used for exon 18 to 30 direct sequencing

exon		5->3
18	F	GCCAGATGAGACTGGCATTT
18	R	ACCCCTCTGGCTGGATTAAC
19	F	GAACATTATAAGACAGTAAATGGAATG
19	R	ACGTGCCTTGCCCTAAGAC
20	F	TTGGGTGGCTGTGTGTTATG
20	R	TTTAAGGTCACCCTCCCTCA
21	F	CAAAATAACATTCCAGAGACCCTA
21	R	CCGATTTCAAACCAAAACTGA
22	F	GGACGCACACACAGACTTCTAC
22	R	ACAATGAATGAGATGCAGTAGCC
23	F	TGCATTTTGTTGGTTTGACAAT
23	R	GGGGACAATTTCTACAAGTTTCTAA
24	F	TGCTGTTGAAGCCCTCTCAC
24	R	CAAGAGCTTTGCAGAGAGCA
25	F	CTGGTGTGCAGAAGCACCT
25	R	CACGGCTCACTGAATGACAC
26	F	TTCCAGGGTGTTGTTTGTTG
26	R	GGATGGAAAAATAAAAACGCATA
27	F	CTTAAAGGCAGGGCCGATT
27	R	TGCAAGAAAAAGGCACACAA
28	F	CACACATGCATGGGACTCTG
28	R	GACACGTGGGCAATGGAG
29	F	ACTTGCCTGGTCTCACAGC
29	R	TGCGAGTCTTTCCCTCCTC
30	F	ACCACTGGAGGTGCCATGT
30	R	ACAGGATGCTTCGTCAGACC

3.2.1.2 Primer preparation

Primers were supplied as ordered by (Eurofins MWG, Germany). Primers were diluted with DNAse Free water to the volume recommended by the manufacturer, to make a final concentration of 100pmol/µl. A dilution of 10pmol/µl for each primer was then prepared ready for use. This was done by transferring 20 µl of the 100pmol/µl primer and adding 180 µl of DNAse Free water.

3.2.1.3 PCR reaction conditions

PCR was carried out using Amplitaq Gold PCR Master Mix (Applied Biosystems, Foster City, CA, USA) as listed in **Table 6** below;

Table 6 PCR reaction mix volumes used for all of the exons to be sequenced. The only difference for each exon was the corresponding forward and reverse primer

Reagent	Volume (µl)
Amplitaq Gold PCR Master Mix	12.5
DNA	2 (= 100ng)
Forward primer	0.5 (at 10pmol/µl)
Reverse primer	0.5 (at 10pmol/µl)
Sterile H_2O	9.5

PCR conditions used in the thermocycler (Applied Biosystems, GeneAmp, PCR System 9700) are listed in the table below

Table 7 PCR conditions, for amplification of the target exon fragment, as set on the thermocycler . The annealing temperature was changed based on the optimum temperature for the primer pair used, specified under the table.

Step	Temperature (°C)	Time (seconds)
Initial denaturation	95	600
	35 Cycles	
Denaturation	95	30
Primer annealing	Depends on primer pair Either 55 or 60	45
Extension	72	45
Final extension	72	600
Hold	4	∞

Annealing temperatures for primer pairs were used as follows;

For exons 18 to 23 = 55°C

For exons 24 to 30 = 60°C

3.2.1.4 PCR product check

PCR products were checked by loading 2.5 µl of the PCR product, with 2.5 µl 1x loading dye (1 µl of 6x loading dye and 5 µl of TBE buffer), onto a 2% TBE agarose gel. The Hi-Lo (Bionexus, Boston, MA, US) marker was used and the gel was run at 80V for 45 minutes. Product bands were compared to the marker corresponding to the expected product size.

3.2.2 PCR product purification

Purification of PCR products was carried out using QIAquick PCR purification kit, manufactured by Qiagen (Hilden, Germany), using the microcentrifuge method. The kit is supplied with the following reagents listed in **Table 8**.

Table 8 showing the reagents supplied with QIAquick kit

Reagent
PB buffer
PE buffer
Elution buffer

When first using the kit, ethanol (96-100%) should be added to the PE buffer. The purification steps are as follows;

1- To the PCR sample, 5 times the volume of buffer PB were added to 1 volume of PCR sample, in most cases it was 100 µl of buffer PB to 20 µl of sample and mixed.

2- QIAquick spin column were placed in the 2ml collection tube supplied with the kit. The mix of buffer PB and PCR sample were added to the column

3- The tubes were spun for 1 min at maximum speed and then the flow-through was discarded.

4- To each QIAquick column, 0.75 ml of Buffer PE was added for washing and the columns were then spun for 1 min at maximum speed and the flow-through was discarded and the columns placed back in the same tube for an additional 1 minute spin. This was done to completely remove any residual ethanol from Buffer PE, which is achieved only if the flow-through is discarded before the additional centrifugation.

5- The QIAquick columns were then placed in clean 1.5ml microcentrifuge tubes. DNA was then eluted by adding 30 µl of elution buffer (10mM Tris.Cl, pH 8.5) to the centre of the QIAquick membrane and the column was left to stand for 1 min before centrifuging for a minute. The purified products were either used directly or stored at -20 °C.

3.2.3 Sequence PCR

3.2.3.1 Sequence PCR reaction mix

Sequence PCR was carried out using ABI PRISM® BigDye™ v3.1 Terminator Cycle Sequencing Kit (Applied Biosystems).

A listing of the kit reagents is given below:

• Ready Reaction Mix

• pGEM®-3Zf(+) double-stranded DNA Control Template

• –21 M13 Control Primer (forward)

• BigDye Terminator v1.1/3.1 Sequencing Buffer (5X)

Aliquots of the Ready Reaction Mix were made to avoid excessive freeze thaw cycles. The mix was prepared for either 5 or 10 samples at a time using the reagent listed in **Table 9** below.

Table 9 the reagent volumes for Sequence PCR reaction using the BIgDye kit

Reagent	Volume (µl)
Nuclease Free H_2O	2
BigDye Terminator v1.1/3.1 Sequencing Buffer	2
Either reverse or forward primer (1.6pmol/µl)	2
Ready Reaction Mix	2
DNA	2

3.2.3.2 Sequence PCR thermocycler program

PCR was carried out using the thermocycler GeneAmp, PCR System 9700 (Applied Biosystems) and set at the conditions shown in the table below.

Table 10 showing the PCR conditions used for the Sequence PCR step

Step	Temperature (°C)	Time (seconds)
Initial denaturation	95	300
	25 Cycles	
Denaturation	95	20
Primer annealing	50	20
Extension	60	240
Hold	4	∞

3.2.4 Purification of sequence PCR

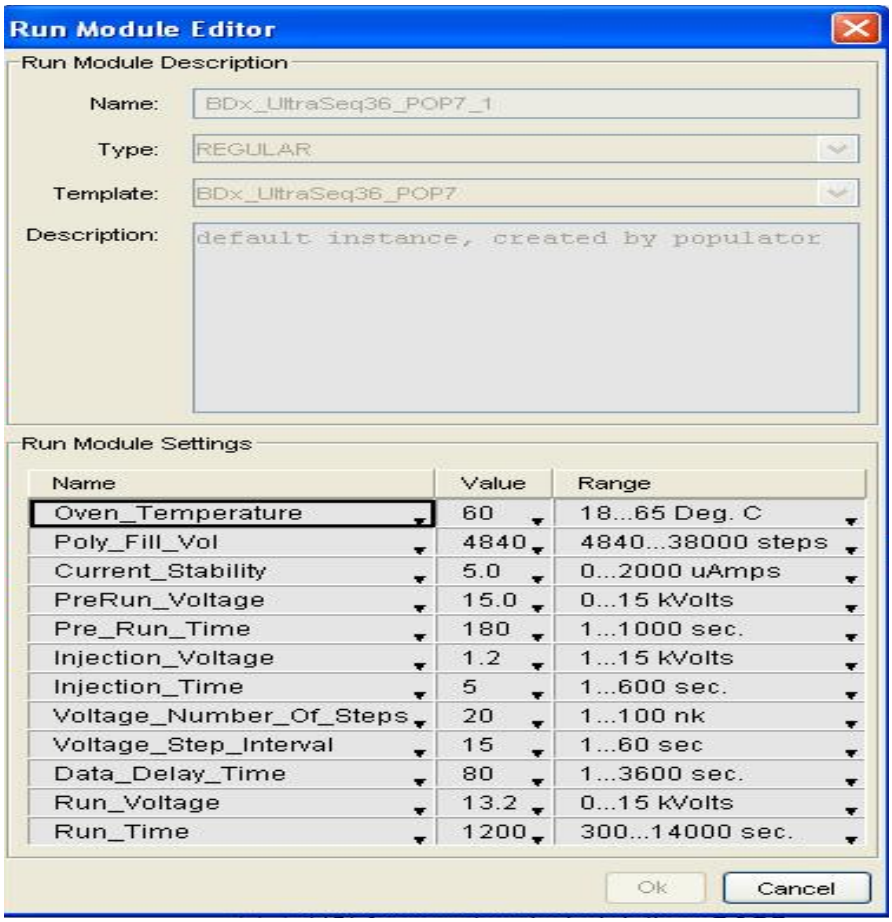
Purification of PCR product was carried out using BigDye Xterminator purification kit (Applied Biosystems). Samples were first transferred onto a PCR plate. To each well 45 µl of SAM solution and 10 µl of BigDye Xterminator solution was added. The plate was covered and attached to the vortex and it was left to shake for 30 minutes. Afterwards, the plate was placed in a centrifuge (Hermle z383k) at a speed of 1000xg for 2 minutes.

3.2.5 loading samples and Sequencer settings

The samples were loaded onto ABI 3130 Genetic Analyzer. Clicking on the Run 3130 Data Collection icon started the machine.

The plate layout was first set on the software to indicate priority (100), Result Group (Sequencing), instrument protocol 1 (B0xUltraSeq_POP7_v3.1) and Analysis protocol (3130Pop7_v3.1). These are the stored settings optimised by the machine suppliers.

Figure 37 Screen shot of the sequencer run module and the values applied



The plate was then loaded onto the machine. The run conditions are illustrated in **Figure 37** above.

The results were analysed by the software and they are presented and discussed further in chapter 4.

3.3. Multiplex Ligation-dependant Probe Amplification

Following the recommendations for ID screening, three Multiplex Ligation-dependant Probe Amplification (MLPA) kits were used supplied by MRC-Holland. These were P245-A2 microdeletions, P036-E1 Telomere-3, and P070-B2 telomere-5. These cover probes that have been reported to cause ID and it is used as a second method to

confirm CNVs found in the array results or as part of an initial screening step for known syndromes.

MLPA kit P245-A2

This kit contains 49 probes for 21 different microdeletion syndromes that cause mental retardation. Their sizes range from 130 to 484 nt. In comparison to previous version of this kit, two additional control fragments at 100 and 105 nt have been included (X and Y-specific), the 108 nt Y probe was removed from the updated kit. It is worth noting that the letter and number following the dash (-) in the kit name indicate the version of the kit.

Other syndrome-specific MLPA kits that have more probes and cover a specific chromosomal region can be used to confirm results generated by this kit. Also, other kits are available to detect changes in methylation pattern. These kits were not used in this study.

These include probe fragments for 2p16.1 microdeletion syndrome, 3q29 microdeletion syndrome, Wolf-Hirschhorn syndrome, Cri du Chat region, Sotos syndrome, Williams syndrome, Langer-Giedlon syndrome, 9q22.3 microdeletion, DiGeorge region 2, WAGR syndrome, Prader-Willi/Angelman syndrome, 15q24 deletion syndrome, Rubinstein-Taybi syndrome, Miller-Dieker region, Smith-Magenis syndrome, NF1 microdeletion syndrome, 17q21.31, DiGeorge syndrome, Phelan-McDermid syndrome, RETT syndrome, and MECP2 duplication syndrome. **Table 11** below lists the MLPA kit P245-A2 microdeletion probmix, inclusive of genes, chromosomal location and related syndrome of each probe.

Table 11 a list of probes in MLPA kit P245-A2 microdeletion probmix

Length (nt)	SALSA MLPA probe	Chromosomal position	Syndrome
64-70-76-82	Q-fragments: DNA quantity; only visible with less than 100 ng sample DNA		
88-92-96	D-fragments: Low signal of 88 or 96 nt fragment indicates incomplete denaturation		
100	X-fragment: Specific for the X chromosome (new from lot 1008 onwards)		
105	Y-fragment: Specific for the Y chromosome (new from lot 1008 onwards)		
118	Y-fragment: Specific for the Y chromosome		
130	**TNFRSF4** Probe 02269–L01761	1p36.33	1p36 deletion syndrome
136	**GATA3** Probe 07632–L07317	10p15.1	DiGeorge region 2 (10p)
142	**PAFAH1B1** Probe 04120–L03532	17p13.3	Miller-Dieker region
148	**MECP2** Probe 09310–L09999	**Xq28**	MECP2 / Xq28 duplication
153	**NSD1** Probe 02595–L08077	5q35.3	Sotos syndrome
160	**UBE3A** Probe 04620–L00863	15q11.2	Prader-Willi / Angelman
166	**GABRD** Probe 04690–L07966	1p36.33	1p36 deletion syndrome
172	**CREBBP** Probe 03087–L02487	16p13.3	Rubinstein-Taybi syndrome
178	**GNB1** Probe 02890–L07968	1p36.33	1p36 deletion syndrome
184	**MECP2** Probe 09311–L10002	**Xq28**	MECP2 / Xq28 duplication
190 *	**SEMA7A** Probe 08380–L10003	15q24.1	"15q24 deletion syndrome"
196	**CLDN5** Probe 01218–L06270	22q11.21	DiGeorge syndrome
202	**MECP2** Probe 03409–L02797	**Xq28**	MECP2 / Xq28 duplication
208	**GP1BB** Probe 05464–L10114	22q11.21	DiGeorge syndrome
214	**NDN** Probe 06282–L01542	15q11.2	Prader-Willi / Angelman
220	**PAX6** Probe 03253–L02690	11p13	WAGR syndrome
226	**MAPT** Probe 07856–L08385	17q21.31	"17q21 microdeletion"
232	**LETM1** Probe 04190–L05920	4p16.3	Wolf-Hirschhorn region
238	**PAFAH1B1** Probe 01443–L08394	17p13.3	Miller-Dieker region
246	**SNRPN** Probe 02026–L10004	15q11.2	Prader-Willi / Angelman
254 †	**SHANK3** Probe 10181-L11409	22q13.33	22q13 / Phelan-McDermid
260	**NF1** Probe 03778–L11180	17q11.2	NF1 microdeletion syndrome
267	**FANCL** Probe 08386–L11411	2p16.1	"2p16.1 deletion syndrome"
274	**LRRC48** Probe 01452–L00936	17p11.2	Smith-Magenis syndrome
283 ±	**CRR9** Probe 01126–L00684	5p15.3	Cri du Chat syndrome
292	**SNRPN** Probe 01318–L07970	15q11.2	Prader-Willi / Angelman
297 †	**DMD** Probe 01412–L01059	**Xp21.2**	Chromosome X control probe
304	**LLGL1** Probe 01453–L08499	17p11.2	Smith-Magenis syndrome
310	**ELN** Probe 01333–L00876	7q11.23	Williams syndrome
319	**TGFBR1** Probe 04652–L04036	9q22.33	"9q22.3 deletion syndrome"
326	**CYP1A1** Probe 06811–L06406	15q24.1	"15q24 deletion syndrome"
335	**NF1** Probe 02508–L02620	17q11.2	NF1 microdeletion syndrome
341	**MAPT** Probe 07857–L08501	17q21.31	"17q21 microdeletion"
349	**Hs.538604** Probe 01232–L07388	10p15.1	DiGeorge region 2 (10p)
355	**DLG1** Probe 08395–L08249	3q29	"3q29 deletion syndrome"
364	**ELN** Probe 01336-L00878	7q11.23	Williams syndrome
373	**SNAP29** Probe 01235–L00773	22q11.21	DiGeorge syndrome
382 †	**SHANK3** Probe 10182–L11174	22q13.33	22q13 / Phelan-McDermid
391	**LIMK1** Probe 01337–L02333	7q11.23	Williams syndrome
401	**TRPS1** Probe 03081–L07411	8q24.12	Langer-Giedion syndrome
409	**TGFBR1** Probe 04653–L10006	9q22.33	"9q22.3 deletion syndrome"
418	**DLG1** Probe 08401–L08255	3q29	"3q29 deletion syndrome"
427 †	**EIF3S3** Probe 01108–L00679	8q24.11	Langer-Giedion syndrome
437	**TERT** Probe 03761–L02477	5p15.33	Cri du Chat syndrome
445	**WHSC1** Probe 06058–L05513	4p16.3	Wolf-Hirschhorn region
454	**NSD1** Probe 02600–L02071	5q35.3	Sotos syndrome
466 †	**RAI1** Probe 09440-L11412	17p11.2	Smith-Magenis syndrome
474	**CRHR1** Probe 07859-L07620	17q21.31	"17q21 microdeletion"
484 †	**REL** Probe 09860-L10628	2p16.1	"2p16.1 deletion syndrome"

MLPA kit P036-E1

Around 3-8% of mental retardation cases are due to copy number variations of subtelomeric regions (Rooms et al 2005; Christofolini et al 2010). This kit contains one MLPA probe for each subtelomeric region therefore totalling 46 probes in addition to

10 control fragments. **Table 12** below lists the MLPA kit P036-E1 microdeletion probmix, inclusive of genes, chromosomal location and related syndrome of each probe.

Table 12 a list of probes in MLPA kit P036-E1 microdeletion probmix

Length (nt)	Chromosomal position	Gene detected	SALSA MLPA probe	MapView build 36 position
64-70-76-82	Q-fragments: DNA quantity; only visible with less than 100 ng sample DNA			
88-92-96	D-fragments: Low signal of 88 or 96 nt fragment indicates incomplete denaturation			
100	X-fragment: Specific for the X chromosome			
105	Y-fragment: Specific for the Y chromosome			
118	Y-fragment: Specific for the Y chromosome			
130 §	1p	TNFRSF4	02269-L01761	01-001.14
136	2p	ACP1	02274-L08758	02-000.25
142	3p	CHL1	01721-L01329	03-000.34
151	4p	PIGG (FLJ20265)	02005-L02047	04-000.50
158	5p	PDCD6	01723-L01327	05-000.37
166	6p	IRF4	01724-L02048	06-000.34
172	7p	ADAP1 (CENTA1)	02275-L02049	07-000.93
179	8p	FBXO25	02397-L01845	08-000.40
186	9p	DMRT1	01727-L02050	09-000.84
193	10p	DIP2C (KIAA0934)	02277-L01768	10-000.48
202	11p	RIC8A (RIC-8)	03315-L02733	11-000.20
208	12p	SLC6A12	02276-L01767	12-000.17
219 +	13q-cen	PSPC1	02399-L01847	13-019.24 (Acrocentric chromosome)
227 +	14q-cen	CCNB1IP1 (HEI10)	01732-L01318	14-019.86 (Acrocentric chromosome)
235 +	15q-cen	MKRN3	07291-L08858	15-021.36 (Acrocentric chromosome)
242	16p	POLR3K	01734-L01316	16-000.04
250	17p	RPH3AL	01735-L01315	17-000.17
258	18p	USP14	01736-L02051	18-000.19
265	19p	CDC34	01737-L01313	19-000.49
274	20p	SOX12	02396-L01844	20-000.26
283 +	21q-cen	RBM11	01739-L01311	21-014.51 (Acrocentric chromosome)
289 +	22q-cen	BID	01740-L01310	22-016.61 (Acrocentric chromosome)
298	Xp/Yp (PAR1)	SHOX	01148-L01331	X/Y-000.52 (PAR1 region)
307	1q	SH3BP5L (KIAA1720)	02392-L02149	01-247.08 (0.2 Mb from telomere)
313	2q	CAPN10	01742-L01308	02-241.18 (1.6 Mb from telomere)
322	3q	BDH1	02013-L02052	03-198.76 (0.7 Mb from telomere)
330 §	4q	TRIML2	12050-L11446	04-189.26 (2.0 Mb from telomere)
337	5q	GNB2L1	03319-L02737	05-180.60 (0.2 Mb from telomere)
346	6q	PSMB1	01746-L01304	06-170.69 (0.5 Mb from telomere)
355	7q	VIPR2	01747-L01303	07-158.60 (0.3 Mb from telomere)
361	8q	ZC3H3 (KIAA0150)	01748-L01302	08-144.69 (1.6 Mb from telomere)
372	9q	EHMT1	08205-L08170	09-139.83 (0.2 Mb from telomere)
379	10q	PAOX (PAO)	09142-L09953	10-135.05 (0.2 Mb from telomere)
386	11q	NCAPD3 (KIAA0056)	01751-L01299	11-133.60 (1.2 Mb from telomere)
395	12q	ZNF10	02687-L02154	12-132.24 (0.2 Mb from telomere)
402	13q	F7	01753-L01297	13-112.82 (1.3 Mb from telomere)
411	14q	MTA1	02778-L02201	14-105.00 (1.3 Mb from telomere)
418	15q	ALDH1A3	01755-L01295	15-099.26 (1.0 Mb from telomere)
426	16q	GAS8 (GAS11)	03201-L02669	16-088.63 (0.2 Mb from telomere)
434	17q	TBCD	01757-L01293	17-078.45 (0.5 Mb from telomere)
441	18q	C18orf22 (FLJ21172)	01758-L01292	18-075.90 (0.2 Mb from telomere)
450	19q	CHMP2A (BC-2)	09143-L10626	19-063.75 (0.9 Mb from telomere)
458	20q	OPRL1	02688-L02884	20-062.19 (0.2 Mb from telomere)
466	21q	PRMT2 (HMT1)	02586-L02059	21-046.89 (0.1 Mb from telomere)
475	22q	RABL2B	01762-L08761	22-049.55 (0.1 Mb from telomere)
483	Xq/Yq (PAR2)	VAMP7 (SYBL1)	01763-L02150	X/Y-154.78 (PAR2; 0.1 Mb from telom.)

MLPA kit P070-B2

Similar to kit P036-E1 this kit contains probes for each subtelomeric region but either on a different gene or exon than that included in the previous kit and hence complements the coverage of the sub-telomeric regions. **Table 13** below lists the MLPA kit P070-B2 microdeletion probmix, inclusive of genes, chromosomal location and related syndrome of each probe.

Table 13 a list of probes in MLPA kit P070-B2 microdeletion probmix

Length (nt)	Chromosomal position	Gene detected	SALSA MLPA probe	MapView build 36 position
64-70-76-82	Q-fragments: DNA quantity; only visible with less than 100 ng sample DNA			
88-92-96	D-fragments: Low signal of 88 or 96 nt fragment indicates incomplete denaturation (autosomal)			
100	X-fragment: Specific for the X chromosome (AMOT gene ; X-111.95)			
105	Y-fragment: Specific for the Y chromosome (UTY gene ; Y-013.98)			
118	Y-fragment: Specific for the Y chromosome (DDX3Y gene ; Y-013.54)			
132	1q	SH3BP5L (KIAA1720)	04084-L03605	01-247.08
139	2q	ATG4B (=APG4B)	02781-L03168	02-242.25
145	3q	KIAA0226	02690-L02842	03-198.88
152	4q	FRG1	02691-L02843	04-191.10
160	5q	GNB2L1	02790-L02232	05-180.60
166	6q	TBP	02694-L02844	06-170.71
172	7q	VIPR2	02793-L03167	07-158.63
179	8q	RECQL4	02695-L00610	08-145.71
186	9q	EHMT1	02792-L02846	09-139.78
193	10q	ECHS1	02696-L02847	10-135.03
202	11q	IGSF9B (=KIAA1030)	02697-L02848	11-133.29
211	12q	ZNF10	02686-L02849	12-132.24
218	13q	CDC16	02698-L00753	13-114.03
226	14q	MTA1	02699-L02850	14-104.99
233	15q	TM2D3 (=FLJ22604)	02701-L02851	15-100.01
241	16q	GAS8 (=GAS11)	02702-L00734	16-088.64
250	17q	SECTM1	02703-L03169	17-077.87
258	18q	CTDP1	02704-L03607	18-075.58
265	19q	CHMP2A (=BC-2)	02705-L02853	19-063.76
274	20q	UCKL1 (=FLJ20517)	02706-L00642	20-062.05
281	21q	S100B	02587-L02854	21-046.85
290	22q	ARSA	02707-L00661	22-049.41
298	X/Yq (PAR2)	VAMP7 (=SYBL1)	02708-L02855	X-154.82 + Y-057.68 (PAR region)
306	1p	TNFRSF18	02270-L01762	01-001.13
315	2p	ACP1	02709-L02856	02-000.27
323	3p	CHL1	02896-L02363	03-000.34
329	4p	PIGG	14440-L16146	04-000.51
337	5p	CCDC127 (=LOC133957)	02791-L02233	05-000.26
346	6p	IRF4	04077-L03462	06-000.34
355	7p	UNC84A	02780-L02857	07-000.84
362	8p	FBXO25	02715-L00973	08-000.40
370	9p	DOCK8 (=FLJ00026)	02716-L00688	09-000.38
379	10p	ZMYND11 (=BS69)	05180-L16343	10-000.22
387	11p	BET1L	02784-L02226	11-000.20
393	12p	JARID1A (=RBBP2)	02787-L02229	12-000.29
402 +	"13p"	PSPC1	02717-L03608	13-019.25 (Acrocentric)
409 +	"14p"	PARP2 (=ADPRTL2)	02718-L00732	14-019.90 (Acrocentric)
418 +	"15p"	NDN	04026-L01542	15-021.48 (Acrocentric)
427	16p	DECR2	02720-L00648	16-000.40
436	17p	RPH3AL	04081-L03465	17-000.18
444	18p	THOC1	02789-L02231	18-000.20
450	19p	PPAP2C	03501-L02880	19-000.23
459	20p	ZCCHC3 (=FLJ22115)	02723-L00641	20-000.23
466 +	"21p"	HSPA13 (=STCH)	02724-L00334	21-014.68 (Acrocentric)
478 +	"22p"	IL17RA	02725-L16344	22-015.96 (Acrocentric)
484	X/Yp (PAR1)	SHOX	03714-L16345	X/Y-000.52 (PAR region)

All three kits were supplied with the same reagents listed on **Table 14**. The steps involved include DNA denaturation, hybridization, ligation, PCR amplification, loading onto the electrophoresis machine.

Table 14 List of supplied reagents with each MLPA kit

SALSA MLPA kit component		Volume	Ingredients
1	SALSA MLPA Buffer (yellow cap)	180 µl	KCl, Tris-HCl, EDTA and PEG-6000. pH 8.5
2	SALSA Ligase-65 (green cap)	115 µl	Glycerol, BRIJ (0.05 %), EDTA, Beta-Mercaptoethanol (0.1 %), KCl, Tris-HCl. pH 7.5, Ligase-65 enzyme (bacterial origin)
3	Ligase Buffer A (transparent cap)	360 µl	NAD (bacterial origin). pH 3.5
4	Ligase Buffer B (white cap)	360 µl	Tris-HCl, non-ionic detergents, $MgCl_2$. pH 8.5
5	SALSA PCR Primer Mix (brown cap)	240 µl	Synthetic oligonucleotides, one of which is fluorescently labeled (FAM, Cy5.0 or other dye, dependent on the capillary electrophoresis instrument used), dNTPs, Tris-HCl, KCl, EDTA, BRIJ (0.04 %). pH 8
6	SALSA Polymerase (orange cap)	65 µl	Glycerol, BRIJ (0.5 %), EDTA, DTT (0.1 %), KCl, Tris-HCl, Polymerase enzyme (bacterial origin). pH 7.5
7	Probemix (black cap)	160 µl	Synthetic oligonucleotides, oligonucleotides purified from bacteria, Tris-HCl, EDTA. pH 8.0

3.3.1- Pre-test set up

3.3.1.1 Sample treatment

DNA was diluted to 25ng/ µl, and a total of 5 µl was used for each reaction. The dilution was made in TE buffer (10 mM Tris-HCl pH 8.2 + 0.1 mM EDTA). Samples selected for comparison must have been extracted by the same method, have similar concentration, have originated from the same tissue and undergone similar storage and treatment condition, as the MLPA method is sensitive to these factors and is a method based on relative differences.

3.3.1.2 Selecting Reference samples

A reference sample of DNA is one that is assumed to have normal copy number and hence a sample with no family history of mental retardation was selected, ideally these should be tested with other methods such as FISH technique to confirm that there were no deletions or duplications of these samples (this was not possible in the scope of this study). The manufacturers suggest using a commercial DNA if a reference sample is not available (Promega Catalogue Order Number G1471 male and G1521 female DNA). A negative control was also prepared to contain TE buffer instead of DNA.

N.B. Ligase-65 master mix and Polymerase master mix were prepared within an hour before use and kept on ice as recommended by the manufacturers.

3.3.1.3 Thermocycler programs

All the programs used on the thermocycler throughout the protocol are listed in the **Table 15** below, using thermocycler Gene Amp, PCR System 9700 (Applied Biosystems).

3.3.2- DNA Denaturation

PCR tubes where labelled and 5 µl of DNA sample (25ng/ µl) was placed in the corresponding tubes and TE buffer was added into the negative control tube. The tubes were then placed into the thermocycler using the program listed in **Table 15**. The tubes were removed after reaching 25°C.

3.3.3- Hybridisation

The MLPA buffer and MLPA probemix were vortexed before use and the hybridization master mix was prepared for each reaction as follows:

- 1.5 µl MLPA buffer (yellow cap)
- 1.5 µl probemix (black cap).

N.B. master mix was prepared for 9 samples at a time therefore, for 8 samples to be tested, so 13.5 µl of each reagent was used in the mix, taking one sample aliquot as a extra in case of pipetting errors.

After vortexing the hybridisation master mix, 3 µl was added to each sample tube and mixed well by pipetting up and down. The samples were then placed on the thermocycler using the conditions listed in table 18. Samples were left at 60°C overnight.

Table 15 thermocycler programs used for MLPA

Step	Temperature (°C)	Time (seconds)
1) DNA denaturation (Day 1)	98	300
	25	pause
2) Hybridisation reaction (Day 1)	95	60
	60	16-20 hrs
3) Ligation reaction (Day 2)	54	pause
	54	900
	98	300
	20	pause
4) PCR reaction (Day 2)		
	35 Cycles	
Denaturation	95	30
Primer annealing	60	30
Extension	72	60
	72	1200 (split over 2 steps of 600 sec each)
	15	Pause

3.3.4- Ligation

The ligation buffers were vortexed before use and a Ligase-65 master mix was prepared for each reaction as follows:

- 25 µl dH2O
- 3 µl Ligase buffer A (transparentcap)
- 3 µl Ligase buffer B (white cap)
- 1 µl Ligase-65 enzyme (green cap).

The reagents were mixed well by pipetting gently up and down and samples were then placed on the thermocycler program as listed in table 18. During the pause period at 54°C, 32 µl of the ligase master mix was added to each reaction tube and mixed by gently pipetting up and down.

The program was then continued for 15 minutes incubation at 54°C. The tubes where then removed at the 20°C hold.

3.3.5 PCR reaction

The SALSA PCR primer was vortexed before use. The polymerase was warmed by hand for 10 seconds

The polymerase master mix was prepared for each reaction as follows:

- 7.5 µl dH2O
- 2 µl SALSA PCR primer mix (brown cap)
- 0.5 µl SALSA Polymerase (orange cap).

The reagents were mixed well by pipetting up and down and stored on ice ready for use.

At room temperature, 10 µl of polymerase mix was added to each tube and mixed by pipetting gently up and down and placed on the thermocycler with the program listed. When the PCR reaction products were stored, the storage box was wrapped in aluminium foil.

3.3.6- Electrophoresis using Beckman GenomeLab GeXP CEQ

The MLPA kits were ordered with the Cy5 label to be used with the beckman GenomeLab electrophoresis machine. The capillaries size selected was 33cm as recommended by the manufacturer's.

The program settings used were as follows:

Initial settings:

- Capillary temperature: 50°C
- Denaturation: 90°C for 120 sec
- Injection voltage: 1.6 kV
- Injection time: 30 sec
- Runtime: 60 minutes at 4.8 kV.

The injection mixture was prepared as follows:

- 0.7 µl PCR reaction
- 0.2 µl Beckman D1-labeled CEQ size standard 600
- 32 µl formamide (or Beckman Sample Loading Solution).

The mixture was prepared in a 96-well plate and one drop of high quality mineral oil was added to each well. The plate was briefly heated before loading the plate onto the electrophoresis machine.

The results were collected and analysed as discussed and presented in chapter 5.

3.4 Cytogenetic assay using Affymetrix 2.7M cytogenetic arrays

The cytogenetic assay introduced by Affymetrix is meant to give higher resolution for the detection of micro deletions or gains that would not be visible under the microscope and hence missed by traditional karyotyping methods. It will also give a view at a whole genomic level.

The steps involved include; amplification of whole genomic DNA, denaturation, fragment amplification, washing and labelling before hybridization onto the array.

Each array contains probes with coverage of 2.7 million markers across the genome, including 400,000 single nucleotide polymorphisms (SNPs)

3.4.1 Whole genome amplification

The reagents supplied for this step are listed in **Table 16** below. DS is the denaturation solution, NS neutralization solution, and Amplif Rxn is the amplification reaction buffer

Table 16 Reagents supplied with the Cytogenetics kit for whole genome amplification

Reagent		
From the Cytogenetics Reagent Kit	Kit Module	Cap Color
▪ Cyto 10X Denat Soln (DS) (Denaturation Solution)	–20 °C	Red
▪ Cyto 10X Neutral Soln (NS) (Neutralization Solution)	–20 °C	Red
▪ Cyto Amplif Rxn Buffer	–20 °C	Red
▪ Cyto Amplif Enzyme Mix	–20 °C	Red
▪ Cyto Water	Room Temperature	Clear

3.4.1.1-Preparation

DNA was measured for all samples and dilution was made with 1x elution buffer (Promega, Leiden, Netherlands) to achieve a final concentration of 33ng/µl. From each sample 3 µl was transferred to appropriately labelled sterile PCR tube.

The thermocycler (Applied Biosystems, GeneAmp, PCR system 9700) was first preheated for 10 min before use at 30ºC.

The Cyto Amplif Rxn Buffer was thawed at room temperature then vortexed vigorously at maximum speed for 10s and immediately placed on ice.

Cyto 10X Denat and Neutral solutions were thawed at room temperature, vortexed and spun, then left at room temperature before dilution, based on manufacturers' recommendations. (Only 1-4 samples were run at a time including control with an exception of one run which involved 5 samples one of which was the control DNA).

3.4.1.2 Diluting the Cyto 10X Denat and Neutral Solutions to 1X

Dilution was carried out as listed below based on the number of samples to be amplified and using the water supplied within the kit. Each buffer was diluted in 1.5 mL

Eppendorf tube, vortexed and spun. After dilution 1X Denat Soln was kept at room temperature, while 1X Neutral Soln was placed on ice.

Table 17 The volumes to add of 10x Denat Soln for the number of samples used

Number of Samples	Volume of Cyto 10X Denat Soln (DS)		Cyto Water		Final Volume of 1X Denaturation Solution (20% extra)
1 to 4	2 µL	+	18 µL	=	20 µL
5 to 8	4 µL	+	36 µL	=	40 µL

Table 18 The volumes to add of 10x Neutral Soln for the number of samples used

Number of Samples	Volume of Cyto 10X Neutral Soln (NS)		Cyto Water		Final Volume of 1X Neutralization Solution (20% extra)
1 to 4	4 µL	+	36 µL	=	40 µL
5 to 8	8 µL	+	72 µL	=	80 µL

(N.B. **Tables 17 and 18** are copied from Cytogenetics Assay user manual based on the volumes required for this study, more volumes are available for more samples)

3.4.1.3. Adding 1X Denat and Neutral Solutions to Samples

To each sample 3 µL of 1x Denat Soln was added by pipetting to the wall of the tube, making a total volume of 6 µL. Samples were then vortexed and spun and incubated at room temperature for 3 min. Tubes where then placed on an aluminum block on ice and *immediately* added 6 µL 1X Neutral Soln by pipetting to the wall of the well, sealed, vortexed and spun before placing the tubes back on ice.

3.4.1.4. Preparing and adding Amplification Master Mix

To mix the Cyto Amplif Enzyme Mix, flicking the tube was followed by a quick spin. The master mix was prepared based on manufacturer's recommendation as listed in ***Table 19*** *below.*

Table 19 Master mix preparation volumes

Reagent	Number of Samples 1	4 (+20%)
Cyto Amplif Rxn Buffer	50 µL	240 µL
Cyto Amplif Enzyme Mix	2.5 µL	12 µL

(N.B. Volumes for more samples are available in the user manual)

The master mix was then vortexed (50% maximum speed), then flicked and inverted the tube 2X. This was repeated before a quick spin and placing the master mix on ice. To each sample 52.5µL of the master mix was added (bringing the total volume to 64.5µL/tube). Samples were then sealed, vortexed and spun, before loading them onto the thermocycler to run the cyto Amp program (as listed in the **Table 20** below)

Table 20 showing the program used on the thermocycler

Cyto Amp Program	
Temperature	**Time**
30 °C	16 hr
65 °C	3 min
4 °C	Hold

After the program was completed and depending on the time available the next stage was carried out or samples were stored at -20 ºC.

3.4.2 Purification & Quantification

The reagents used are listed in **Table 21** below.

Table 21 showing the reagents supplied for the purification stage

Reagent		
From the Cytogenetics Reagent Kit	Kit Module	Cap Color
▪ Cyto Elution Buffer	Room temperature	Green
▪ Cyto Magnetic Beads	4 °C	Green
▪ Cyto Purif Wash Buffer	Room temperature	Green

3.4.2.1. Adding Magnetic beads to amplified samples

The Cyto Magnetic bead bottle was inverted ten times to thoroughly mix until the solution appeared homogenous with no bead clumps on the bottom of bottle. Samples were vortexed and spun and then 90µL of Cyto Magnetic beads were added to each sample and slowly pipetting up and down 10X to mix before incubating for 5 minutes at room temperature.

The full volume was transferred to a 1.5mL Eppendorf tube, which is then placed on a magnetic-ring stand, covered, and incubated at room temperature for 5 minutes. After setting the pipette to 160µL the supernatant was removed and discarded. As a cautionary measure not to disturb the beads, the pipette tip was placed in the center of bead ring and pulled up after aspiration.

3.4.2.2. Preparing and adding Puri Wash Buffer

For the first use of the kit, 10.5mL of 99.9% ethanol was added to Cyto Puri Wash Buffer bottle for the 24 samples reagent kit. The bottle was inverted 5 times to mix. While the tubes are still placed on the magnetic stand, then 200µL was added to each sample, incubated at room temperature for 1 min and the supernatant was removed and discarded. This was repeated until all the supernatant was removed.

3.4.2.3. Eluting samples

To elute samples, 50µL of Cyto Elution Buffer was added immediately to each sample to avoid drying of the pellets. The tubes were kept at the stand the whole time up to elution and great care was taken not to disturb the pellet. The tubes were then secured on the foam adaptor with tape and vortex at maximum settings for 15 minutes. The tubes were then placed on magnetic stand and incubated at room temperature for 30 minutes. Then 45µL, of the elute from each sample, was transferred to a fresh tube.

3.4.2.4. Quantification

The eluted samples were quantified using the NanoDrop spectrometer (Thermo Scientific) following the software instructions and using Elution buffer for the reference blank sample. A volume of 2µL was used of each sample for quantification. Depending on the time available the samples were either stored at -20ºC or prepared for fragmentation and labelling.

3.4.3-Fragmentation & Labelling

Table 22 The reagents used for fragmentation

Reagent		
From the Cytogenetics Reagent Kit	Kit Module	Cap Color
▪ Cyto Frag and Label Buffer	–20 °C	Blue
▪ Cyto Frag and Label Enz Mix	–20 °C	Blue

3.4.3.1. Preparation and fragmentation

The thermocycler lid was preheated for 10 minutes. Of each sample with an acceptable OD range (concentration greater than 0.55 µg/µL) 37µL was transferred to a fresh tube. The Cyto Frag and Label Buffer was thawed at room temperature and then placed on ice. The tubes were placed on an aluminum block on ice. The Fragmentation and Labelling Master Mix was prepared on ice, as shown in the **Table 23** below.

Table 23 The preparation volumes for the fragmentation step

Reagent	Number of Samples 1	4 (+20%)
Cyto Frag and Label Buffer	10 µL	48 µL
Cyto Frag and Labelling Enz Mix	3 µL	14.4 µL

To each sample 13µl of master mix was added (bringing the total to 50µl). The sample was then vortexed and spun and loaded onto the thermocycler (Applied Biosystems, GeneAmp, PCR system 9700) with the program listed in **Table 24** below.

Table 24 Thermocycler steps for the fragmentation step

Cyto Frag-Label	
Temp	Time
37°C	2 hr
95°C	10 min
4°C	Hold

3.4.3.2. Fragmentation Quality Control gel (optional step)

After the program was completed, the samples were placed in an aluminium block on ice and 2.5 µl of each sample was mixed with 2.5 µl of 1x gel loading buffer (diluted from 6x with 1x TBE buffer, Invetrogen). Samples were then loaded onto 4% TBE agarose gel. As a marker, 5µl of Bionexus Hi-Lo ladder was loaded onto the gel, which was then run at 80V for 40-60 min. Remaining samples were stored at -20°C unless hybridization was followed directly.

3.4.4 Hybridization

Table 25 Reagent supplied for the hybridization step

Reagent		
From the Cytogenetics Reagent Kit	Kit Module	Cap Color
▪ Cyto Hyb Buffer (in a bottle)	–20 °C	Amber

3.4.4.1. Array preparation

The arrays were unpackaged and left to reach room temperature. Each array was labelled and then a 200µl pipette tip was inserted into the upper right septum, as illustrated in **Figure 38** below.

Figure 38 Preparation of array chips for sample loading with tips placed at the to septa and tough spots placed on the top ready to seal the septa after sample loading

The hybridization oven was preheated at 50°C with the rotation stabilized at 60 rpm. The Cyto Hyb Buffer was thawed by placing it on ice for 15 minutes. Samples were thawed if frozen, vortexed and spun, then placed in an aluminum block on ice.

3.4.4.2. Sample denaturation

The Cyto Hyb Buffer bottle was slowly inverted 5 times to mix, of which 100µl was then added to each sample, sealed, vortexed and spun. The samples were then loaded onto the thermocycler and run the Cyto Denature program specified in **Table 26** below;

Table 26 Thermocycler steps for denaturation step

Cyto Denature	
Temp	Time
95°C	5 min
50°C	15 min
50°C	Hold

3.4.4.3. Loading samples onto the arrays and hybridization

The samples were kept on the thermocycler and 130µl of each sample was loaded onto the appropriately labelled, Cytogenetics 2.7M, array using a P200 pipet to inject into the lower left septum of the array. Access fluid around the septa was cleaned using lens-cleaning tissues and tough-spots were firmly applied to the septa. Then the arrays were loaded into the preheated hybridization oven and incubated overnight for 16 to 19 hours at 50°C and a rotations of 60 rpm.

3.4.5 Washing, staining and scanning

Table 27 Reagents supplied for the washing and staining

Reagent		
From the Cytogenetics Reagent Kit	Kit Module	Cap Color
▪ Cyto Stain Buffer 1	4 °C	Amber
▪ Cyto Stain Buffer 2	4 °C	Amber
▪ Cyto Holding Buffer	4 °C	Amber
▪ Cyto Wash Buffer A	Room temperature	White opaque
▪ Cyto Wash Buffer B	Room temperature	White opaque

3.4.5.1 Preparing the fluidics stations and priming

The fluidics station (GeneChip® Fluidics Station 450) were prepare by making sure water was loaded in the appropriate bottle and suction tubes were connected to the appropriate bottles, Cyto Wash buffer A, Cyto Wash buffer B and a waste bottle.

Affymetrix launcher was started to initiate the fluidics control software module before switching the fluidics stations on. Affymetrix GeneChip® Command Console (AGCC) is used to control the fluidics station and the scanner.

PRIME_450 protocol was selected and copied to all four stations, then, the Copy to Selected Modules option was clicked before selecting the run all for the protocol. Instructions were followed to replace the Eppendorf tubes under the needles in each station as prompted by the load three empty vials command. The tubes were loaded and locked in place for priming to begin. Meanwhile, the arrays were registered through the AGCC portal registration forms. Sample name was generated for each array and the code bar was scanned to complete the registration process.

3.4.5.2 Washing and staining with Cytogenetics_Array_450 protocol

Two tinted and one clear 1.5ml Eppendorf tubes were labelled 1 to 3 respectively for each array. The stains and holding buffer were mixed before use by inverting each bottle 5 times. To tube number 1, 500µl of Cyto Stain Buffer 1 buffer was added, while 500µl of Cyto Stain Buffer 2 was added to tube number 2, and finally 800µl of Cyto Holding Buffer was added to tube number 3. The reagents were then loaded onto the fluidics station with tubes 1 to 3 loaded onto position 1 to 3 respectively.

The Tough-Spots were then removed from the arrays. The Hyb Buffer was removed from the arrays and replaced with Cyto Holding Buffer. The Cytogenetics_Array_450 protocol was then selected copied to the stations to be used and then Run option was

clicked. When prompted to Eject Washblock, the cartridge lever was depressed and then the array was placed in the fluidics station when the Load cartridge prompt appeared. The array was secured into place.

3.4.5.3 Array scan

The GeneChip® Scanner 3000Dx was turned on to allow it to warm up before scanning the arrays. The AGCC Scan control was then selected.

After the wash and stain protocol was completed the arrays were ejected as prompted and checked for any air bubbles. Any array with bubbles was returned to the fluidics station where de-bubbling would automatically take place. The washblocks were then re-engaged for the protocol to complete. The reagent vials were then removed and fresh tubes were placed in the station positions.

Tough-Spots (round seals shown as green circles on **Figure 38**) were placed on the array septa and the array window was wiped with lens paper before being loaded onto the Autoloader at the starting position indicated and then the scanner was sealed and the scan was selected to start and to allow for rescans. When scanning was complete the arrays were removed and stored at 4°C. The protocol shut down was selected for all the fluidics stations before being copied to run.

The results were exported, as .CEL files for analysis with the chromosomal Suite software detailed in Chapter 6, where the results are also presented.

3.5 Sample analysis

Samples collected from the two sets (ID and RTS) and controls were used in different experimental methods and analysis as reagents and supplies were limited to be applied

to all samples. From 27 families 60 samples were collected but only 40 samples were used. Including to three unrelated healthy control samples used for the MLPA method. In **Table 28** below a summary of the samples, affected status, gender and labels for methods used are listed.

Table 28 list of samples used and labels in the applicable methods

Sample #	affected status	gender	Direct sequencing	MLPA (P036-E1/P070-B2)	cytoarrays	exome seq
1.1	**y**	m		1.1	1.1	
1.2	**y**	m			1.2	
2.1	**y**	m		2.1	2.1	
2.2	**y**	f			2.2	
3.1	**y**	m		3.1	3.1	
3.2	**y**	m			3.2	
4.1	**y**	m		4.1	4.1	
5.1	**y**	f		5.1	5.1	
6.1	**y**	f		6.1	6.1	
7.1	**y**	f		7.1	7.1	S0260
7.2	**n**	f			7.2	
7.3	**y**	f			7.3	S0261
8.1	**y**	f		8.1	8.1	
9.1	**n**	m		9.1	9.1	
9.2	**y**	m			9.2	
9.3	**n**	m			failed	
9.3	**n**	m			failed	
10.1	**y**	m			10.1	
11.1	**y**	f		11.1	failed	
12.1	**y**	m			failed	
13.1	**y**	f		13.1		
15.1	**y**	f		15.1		
16.1	**y**	m		16.1		
A.1	**y**	f	7		A.1	
A.2	**y**	m	8		A.2	
B.1	**y**	f	6	B.1	failed	
B.2	**y**	m			failed	
B.3	**n**	f	1			
C.1	y	f	2	C.1		
D.1	y	m	3	D.1		
E.1	y	f	4	E.1		
F.1	y	m	5			
G.1	y	m	9	G.1		
H.1	y	m	10			
I.1	y	f	11	I.1		
J.1	y	f	13	J.1		
J.2	y	f	12			
cc1	n	M		cc1		
cc3	n	M		cc3		
cc8	n	F		cc8		

Chapter 4- CREBBP direct Gene sequencing

4.1 Background

Rubinstein Taybi syndrome (RTS) is a rare genetic disorder affecting 1 in 100,000 newborns (Roelfsema and Peters 2007). Two genes causing RTS have been identified so far, CREBBP (CREB binding protein gene), and its homolog EP300 (E1A binding protein p300). Mutations in CREBBP are thought to account for around 60% of RTS cases, whereby another 3% of cases are caused by mutations in EP300 (Roelfsema and Peters 2007).

In this chapter I will review the current understanding of CREBBP function and reported mutations associated with RTS, and identify potentially novel or reported mutations in samples of RTS cases via direct sequencing of CREBBP exons 18 to 30, where most mutations have been reported to date and the exons cover the HAT domain thought to be vital for the protein's main function.

4.1.1 Role and Function of CREBBP gene

The search for the RTS causing gene started with Petrij and colleagues identifying a chromosomal breakpoint, using FISH technology, at chromosome 16p13.3 in a region of 150kb that covered the CREBBP gene. It codes for a 2442 amino acid protein with a molecular mass of 265kD and encodes a protein with 95% homology with the mouse version. The mode of inheritance of RTS is usually autosomal dominant through *de novo* mutations, where a loss of one gene copy leads to the disorder, suggesting a haploinsufficiency mechanism where one copy of the gene is insufficient for normal function due to dosage sensitivity (Roelfsema and Peters 2007).

CREBBP (also called CREBBP; OMIM 600140) encodes the CREB (cAMP-response-element binding protein) binding protein, which is involved in the transcriptional coactivation of a number of transcription factors that activates cAMP-responsive genes. This has implications on different biological processes such as embryonic development, growth control, and homeostasis, as well as neurodevelopment, synaptic plasticity, and neuroprotection (Sakamoto, Karelina et al. 2011).

The gene expression regulatory function of CREBBP is through the histone acetyltransferase (HATs) domain, which leads to decondensation of the chromatin at specific site for targeted gene transcription. Besides acetylation of histones, CREBBP also acetylates other proteins such as p53.

The CREBBP gene is also associated with another disorder, which is acute myeloid leukemia (Petrij, Giles et al. 1995). Animal studies on CREBBP +/- mouse models, showed that these strains were at higher risk of developing cancerous tumours than their normal counterparts, suggesting a possible role of CREBBP in the tumour suppressor process (Kung et al 2000).

4.1.2 Genetic structure

As illustrated in the **Figure 39** below, CREBBP gene is composed of 31 exons that include key functional domains such as the HAT domain, zinc finger domains, a CREB-binding domain (KIX domain), a bromo domain, and a transactivation domain, which illustrates the diverse function of the gene.

The HAT domain targets histone H3 and H4, and acetylation of these opens the chromatin structure to allow open access for gene expression at that site.

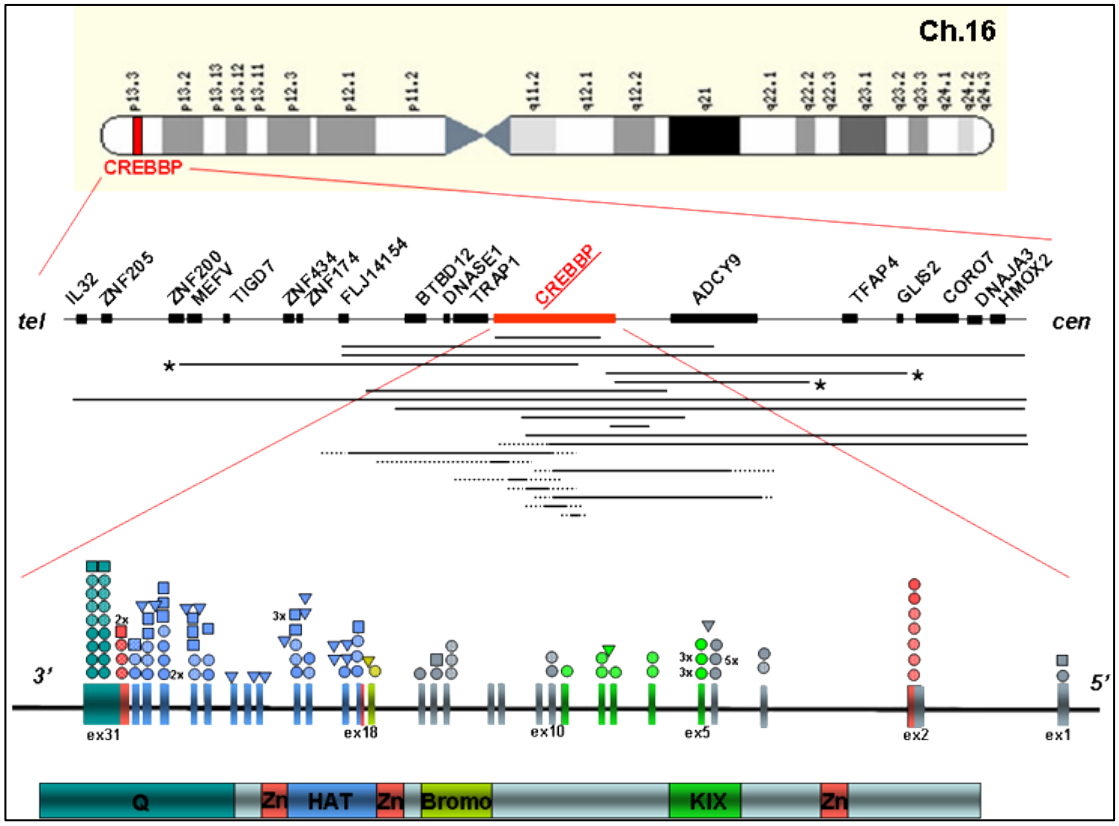


Figure 39 the structure and reported mutations found at different exons

4.1.3 Reported mutations

There are many CREBBP and EP300 mutations reported that have been linked to RTS and these are usually de novo and no hot spots have been identified other than a breakpoint at exon 2 and the mutations at the HAT domain, where malfunction is sufficient to cause the disorder. A list of the reported variations can be found in a paper by Roelfsema and Peters 2007, which were used in this study to compare the variations detected, these are listed in **Table 29** below. Very few publications have reported other mutations since.

Table 29 nonesense mutations, small deletions and insertions, missense, and splice site mutations reported by other studies.

CREBBP nonsense mutations

Exon	Mutation	Protein
Exon 1	c.68C>A	p.S23X
Exon 2	c.304C>T	p.Q102X
Exon 2	c.406C>T	p.Q136X
Exon 4	c.1066C>T	p.Q356X
Exon 4	c.1108C>T	p.R370X
Exon 4	c.1108C>T	p.R370X
Exon 4	c.1108C>T	p.R370X
Exon 4	c.1108C>T	p.R370X
Exon 4	c.1108C>T	p.R370X
Exon 5	c.1237C>T	p.R413X
Exon 5	c.1237C>T	p.R413X
Exon 5	c.1270C>T	p.R424X
Exon 5	c.1270C>T	p.R424X
Exon 5	c.1270C>T	p.R424X
Exon 10	c.1984C>T	p.Q662X
Exon 15	c.2986G>T	p.E996X
Exon 18	c.3517C>T	p.R1173X
Exon 19	c.3639C>A	p.C1213X
Exon 21	c.3805A>T	p.K1269X
Exon 27	c.4398T>A	p.Y1466X
Exon 27	c.4435G>T	p.G1479X
Exon 27	c.4492C>T	p.R1498X
Exon 27	c.4492C>T	p.R1498X
Exon 28	c.4669C>T	p.Q1558X
Exon 29	c.4879A>T	p.K1627X
Exon 31	c.5635C>T	p.Q1879X
Exon 31	c.6010C>T	p.R2004X
Exon 31	c.6019C>T	p.Q2007X
Exon 31	c.6127C>T	p.Q2043X
Exon 31	c.6133C>T	p.Q2045X
Exon 31	c.6283C>T	p.Q2095X

CREBBP Missense mutations

Exon	Mutation	Protein
Exon 1	c.40A>G	p.R14G
Exon 15	c.2941G>A	p.A981T
Exon 18	c.3524A>G	p.Y1175C
Exon 21	c.3832G>A	p.E1278K
Exon 21	c.3833A>G	p.E1278G
Exon 25	c.4238A>C	p.H1413P
Exon 26	c.4304A>T	p.D1435V
Exon 26	c.4340C>T	p.T1447I
Exon 26	c.4348T>C	p.Y1450H
Exon 27	c.4409A>G	p.H1470R
Exon 27	c.4445A>G	p.Y1482C
Exon 28	c.4627G>T	p.D1543Y
Exon 30	c.4951G>C	p.D1651H
Exon 30	c.4991G>A	p.R1664H
Exon 30	c.4991G>A	p.R1664H
Exon 31	c.6661A>C	p.M2221L
Exon 31	c.6728C>T	p.A2243V

Adapted from Roelfsema and Peters 2007

CREBBP deletions and insertions

Exon	Mutation	Protein
Exon 2	c.86_233 del 148	p.D29fsX37
Exon 2	c.139del A ins 13	p.N47delinsIHHEL
Exon 2	c.235 del G	p.G79fsX86
Exon 2	c.474_493 del 20	p.Q158fsX173
Exon 3	c.840 dup T	p.S81fsX281
Exon 3	c.904_905 del AG	p.S302fsX348
Exon 6	c.1381_1388 del 8	p.G461fsX469
Exon 6	c.1481 dup A	p.N494fsX527
Exon 8	c.1733 del C	p.P578fsX588
Exon 8	c.1738 dup A	p.A581fsX586
Exon 9	c.1891_1895 del 4	p.A631fsX632
Exon 10	c.2045 dup A	p.P683fsX725
Exon 14	c.2749 dup A	p.T917fsX969
Exon 14	c.2827 del C	p.Q943fsX997
Exon 16	c.3096 dup T	p.K1033fsX1033
Exon 17	c.3352 dup CC	p.Q1118fsX1130
Exon 18	c.3396_3400 del 5	p.P1132fsX1166
Exon 18	c.3432_3433 del AG	p.T1144fsX1168
Exon 18	c.3545 dup C	p.A1182fsX1186
Exon 20	c.3715_3716 del AA	p.K1239fsX1252
Exon 20	c.3767_3769 del CAC	p.S1256X
Exon 21	c.3824 dup T	p.F1275fsX1282
Exon 25	c.4256_4258 del CT	p.S1419fsX1419
Exon 26	c.4319_4320 del TC	p.F1440fsX1451
Exon 26	c.4322 dup C	p.R1441fsX1452
Exon 27	c.4398_4399insT	p.V1467fsX1467
Exon 28	c.4611 del G	p.L1537fsX1549
Exon 29	c.4837 del G	p.V1613fsX1634
Exon 29	c.4898_4908 del 11	p.F1633fsX1655
Exon 30	c.4945 del A	p.I1649fsX1743
Exon 30	c.4963 del C	p.L1655fsX1743
Exon 31	c.5212 ins 14	p.H1738fsX1748
Exon 31	c.5222 del AG	p.K1741fsX1749
Exon 31	c.5793 dup C	p.1932TfsX1965
Exon 31	c.6043 del A	p.S2015fsX2039
Exon 31	c.6050_6051ins 8	p.P2017fsX2342
Exon 31	c.6065_6071 del 7	p.Q2022fsX2037

CREBBP Splice site mutations

Exon	Mutation	Protein
Exon 4	c.1216 + 1G>A	
Exon 7	c.1676 + 1G>A	
Exon 17	c.3368_3369 + 6 del 7 ins CA	
Exon 19	c.3698 + 1G>A	
Exon 19	c.3698 + 5A>T	
Exon 20	c.3779 + 5G>C	
Exon 21	c.3836 + 1G>A	
Exon 22	c.3837-2 A>T	
Exon 23	c.3915-1 G>A	
Exon 24	c.4133 + 1G>A	
Exon 25	c.4280 + 2T>C	
Exon 27	c.4559A>G	p.1520R/splice site
Exon 28	c.4561-5 C>G	
Exon 28	c.4728 + 1G>A	

4.2 Materials and methods

Laboratory procedure was carried out as described in Chapter 3 using the selected primers and Big dye kit specified.

The primer sequences, reaction mix and PCR conditions are all listed below and further details can be found in chapter 3.

4.2.1 Samples used

At least one sample per family was used for sequencing and all were affected with RTS except for sample 1 corresponding to individual B.3. Details of sample collection and phenotypes can be found in chapter 2 and a summary of the samples used for sequencing is listed in **Table 30** below.

Table 30 showing the samples used in reference to the family number and the label number used for sequencing.

Family number	Individual number	affected status	CREBBP seq number
A (RTS 1)	**A.1**	**Y**	7
A (RTS 1)	**A.2**	**Y**	8
B RTS 2	**B.1**	**Y**	6
B RTS 2	**B.3**	**N**	1
C	**C.1**	**Y**	2
D	**D.1**	**Y**	3
E	**E.1**	**Y**	4
F	**F.1**	**Y**	5
G	**G.1**	**Y**	9
H	**H.1**	**Y**	10
I	**I.1**	**Y**	11
J	**J.1**	**Y**	13
J	**J.2**	**Y**	12

4.2.2 Genetic Analysis

The sequence results are generated in two parts as a pdf file. The first contains an electropherogram that shows the peaks detected by the sequencer and the second contains the resulting sequence as per base calling of the software. Standard interpretation of electropherogram is followed as exemplified by the DNA sequencing Core guidelines created by the University of Michigan available from their website, with examples of a good quality and bad quality electropherogram figures below (http://seqcore.brcf.med.umich.edu/doc/dnaseq/interpret.html).

First the quality of the electropherogram is assessed, below is an example of a good quality electropherogram as the peaks are clear and distinct, with no background noise and the base calling sequence is evenly spaced.

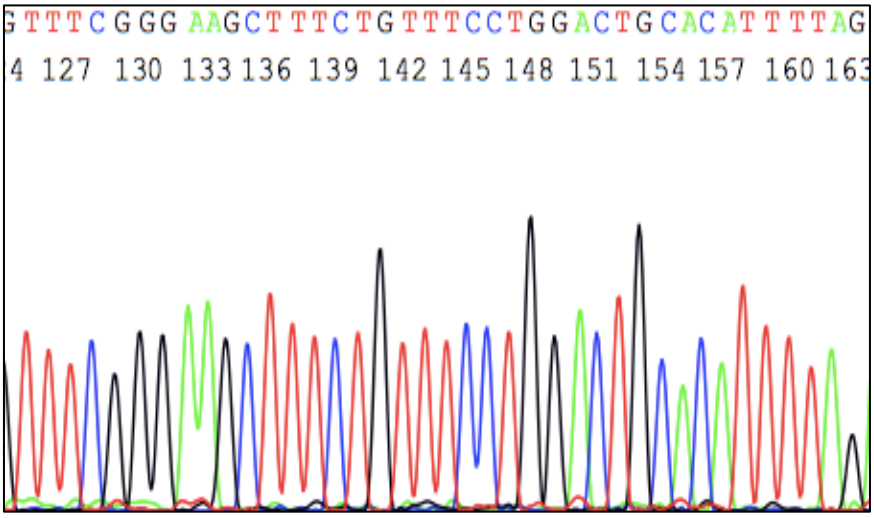


Figure 40 a snapshot of good quality electropherogram results generated by the ABI software provided with the sequencer

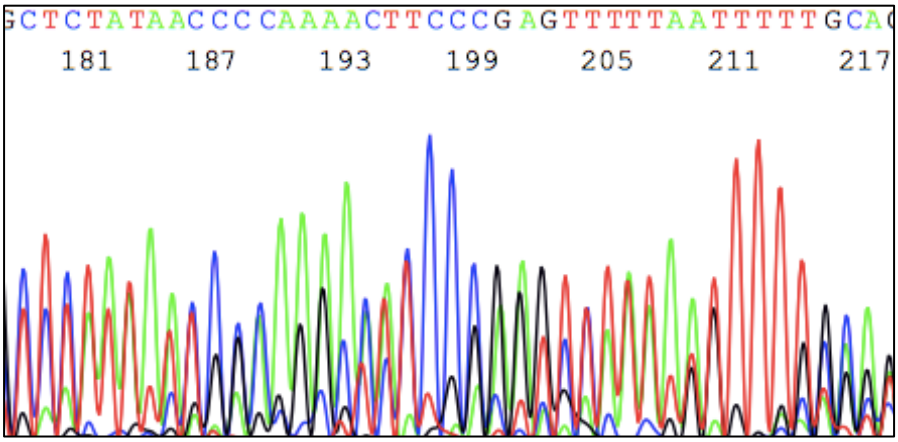


Figure 41 electropherogram of a poor sample with incorrect base calling

While Figure 41 illustrates an example of a poor quality electropherogram that could lead to incorrect base calling. There are a number of other examples of poor quality results illustrated by large spikes of multiple colours within the sequence, which are usually caused by micro-air bubbles or small debris. Another problem that affects results is slippage where DNA separating and re-annealing incorrectly resulting in the disruption of the sequence.

The sequence generated was then copied and entered on the blat tool on the UCSC genome browser (www.genome.ucsc.edu) (Kent 2002). This then generated a sequence alignment with details of percentage of aligned sequence match and starts and end sequence. The results were pooled in the tables shown in the results section. Sequences that did not have a 100% match were looked at further and compared to the coding sequence based on the NCBI ref entry: NM_004380.2 for the 10197bp mRNA sequence for CREBBP. The coding sequence starts with base number 204 where the first open frame is found.

For sequences that were in the reverse order an online reversing tool was used to reverse the sequence for easy comparison (http://www.cellbiol.com/scripts/complement/dna_sequence_reverse_complement.php)

Once there was a variation in the alignment within the coding region this sequence was further checked with an online sequence translation tool (web.expasy.org/translate/) to observe the potential effect on the amino acid sequence.

4.3 Results

Sequencing was carried out for 13 exons on 13 RTS samples with total runs of 169 out of which 9 variations were identified, 67 runs showing perfect alignments and, 43 runs with no variations in the coding region and 50 runs that failed or produced poor quality results.

The results presented here for each exon is of a summary table of all the sample blat results and an alignment of the expected product based of the sequence generated using the e PCR tool (Kent, Sugnet et al. 2002), available at the UCSC website, for each primer set.

The selected result per sample, as exemplified here using exon 18 results in **Figure 42,** was one with the highest alignment score and that matched the targeted sequence. The table columns include the sample number, score (an alignment score generated by the blat search engine that includes a penalty for gaps within the alignment and any base mistmatch), the start end of the alignment point based on the base number of the PCR product sequence, the query sample base size, percentage of identity match within the alignment region, chromosome number, the strand sequence type which is the reverse strand in the case of our sequence, the genomic start and end point, and finally the span of the alignment. The first row represents the results of the expected product using the used primer set for the exon in question. The rows in green indicate that there were no results for the sample for this exon, usually due to sequencing result failure.

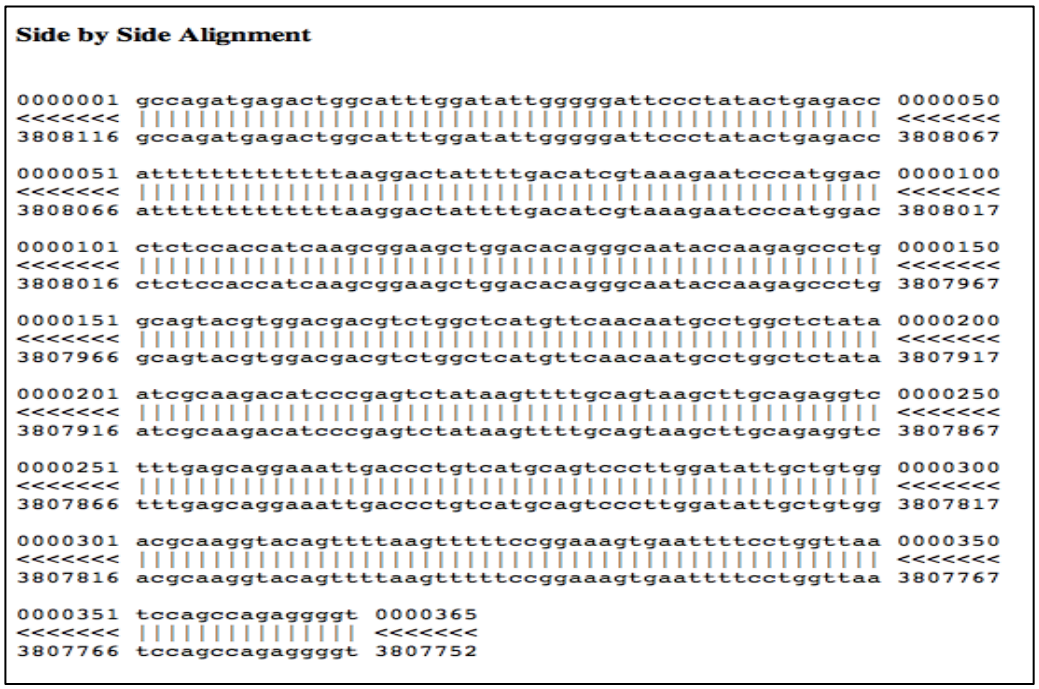
The variations were called using the standard nomenclature, whereby the letter c at the start of the variation code is used to refer to the coding sequence, followed by the position number of the variation and then the variation type (> indicating a substitution while dup is short for duplication del is for deletion and ins for an insertion) with the specific base variation mentioned. In the case of a variation in the amino acid sequence the annotation for the variation starts with the letter p and then followed by the expected amino acid then the position and finally the resulting amino acid.

Direct sequencing results for each exon consisting of a summary table of the blat alignment result for each sample which includes the alignment score, start and end point, query size (Qsize), percentage identity, chromosome, strand (forward or reverse), chromosomal start and end location and alignment span. Missing data (due to sample failure) were highlighted in green. This is followed by a perfect alignment image for the expected fragment. This is exemplified by exon 18 results below. Results for other exons are attached at the appendix.

	Exon 18 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 18	365	1	365	365	100.00%	16	-	3807752	3808116	365
18_1	196	11	274	369	87.50%	16	-	3807827	3808092	266
18_2	38	10	48	370	100.00%	16	-	3808052	3808091	40
18_3	329	1	336	360	99.20%	16	-	3807774	3808111	338
18_4	311	6	325	425	99.10%	16	-	3807780	3808103	324
18_5										
18_6	325	6	341	359	98.60%	16	-	3807780	3808117	338
18_7										
18_8										
18_9	322	6	337	497	98.80%	16	-	3807780	3808115	336
18_10	318	3	330	358	98.80%	16	-	3807778	3808109	332
18_11	329	2	380	416	98.60%	16	-	3807778	3808331	554
18_12	313	6	333	360	97.90%	16	-	3807780	3808109	330
18_13	316	2	333	358	97.90%	16	-	3807775	3808109	335

Figure 42 showing the summary Blat results of PCR products for the RTS samples using exon 18 primers

Figure 42 above shows the Blat results that corresponded with the region in question and represented the highest score match for the alignment. Some fragments showed some alignment results for other regions. The side by side alignment image of the expected product (as produced via the e PCR tool on the ucsc website) is included in **Figure 43** and was used to compare the results for the samples run. Other results are the sequence fragment for each sample, the related electropherogram, and individual alignments for each sample.



Side by Side Alignment

```
0000001 gccagatgagactggcatttggatattgggggattccctatactgagacc 0000050
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3808116 gccagatgagactggcatttggatattgggggattccctatactgagacc 3808067

0000051 atttttttttttttaaggactattttgacatcgtaaagaatcccatggac 0000100
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3808066 atttttttttttttaaggactattttgacatcgtaaagaatcccatggac 3808017

0000101 ctctccaccatcaagcggaagctggacacagggcaataccaagagccctg 0000150
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3808016 ctctccaccatcaagcggaagctggacacagggcaataccaagagccctg 3807967

0000151 gcagtacgtggacgacgtctggctcatgttcaacaatgcctggctctata 0000200
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807966 gcagtacgtggacgacgtctggctcatgttcaacaatgcctggctctata 3807917

0000201 atcgcaagacatcccgagtctataagttttgcagtaagcttgcagaggtc 0000250
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807916 atcgcaagacatcccgagtctataagttttgcagtaagcttgcagaggtc 3807867

0000251 tttgagcaggaaattgaccctgtcatgcagtcccttggatattgctgtgg 0000300
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807866 tttgagcaggaaattgaccctgtcatgcagtcccttggatattgctgtgg 3807817

0000301 acgcaaggtacagttttaagtttttccggaaagtgaattttcctggttaa 0000350
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807816 acgcaaggtacagttttaagtttttccggaaagtgaattttcctggttaa 3807767

0000351 tccagccagaggggt 0000365
<<<<<<< ||||||||||||||| <<<<<<<
3807766 tccagccagaggggt 3807752
```

Figure 43 showing the alignment result for exon 18 using the expected PCR product for the used primer set.

Looking at the tables above for exon 18 results, three samples (18_5, 18_7, and 18_8) failed to yield results. Meanwhile, results of sample 18_2 produced a much shorter fragment than expected, which would require a rerun to confirm that there was no error during sequencing. The electropherogram data (data not shown) for sample 18_1 and 18_2 are of poor quality as peaks are not distinct which would affect the base calling leading to an inaccurate sequence. On the other hand, results of sample 18_3 shows a deletion of two bases that lie outside of the coding region and the same was

true with results of the other samples (18_4, 18_6, 18_9, 18_10, 18_11, 18_12, and 18_13) and some shared the same variation, however, all were outside of the coding region.

For exon 19 results, samples 19_5 and 19_6 failed to yield any sequencing results. However, for sample 19_1, one base insertion was found but outside of the coding region, a closer look at the electropherogram, the peak was not clear and hence it could be a wrong base calling that resulted in the insertion. Results for samples 19_2, 19_3, 19_4, 19_7, 19_10, 19_11, and 19_12 all show perfect alignments and sequence identity match. The results of the remaining samples (19_8, 19_9, and 19_13) had electropherograms of poor quality affecting the base calling and the sequence of the fragment produced leading to fragmented results when aligned.

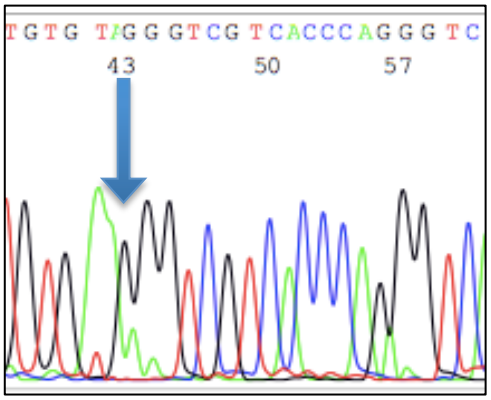


Figure 44: part of an electropherogram for the sequence of exon 20 in individual 9 with the variation indicated by an arrow

For exon 20 results, five samples (20_2, 20_3, 20_ 4, 20_6, and 20_8) showed perfect alignment and sequence match, while sample 20_1 had a variation outside of the coding region. For samples 20_5, 20_7, 20_10, 20_11, 20_12, and 20_13, the electropherogram result was of poor quality in parts and hence affecting base calling and the output sequence. Finally for sample 20_9 result there was a base change c.3766 T>A this leads to an amino acid change p.Ser1255Thr from serine to threonine and this variation was not reported, if it is pathogenic or not is yet to be determined as

only one amino acid is changed. Also, a closer look at the electropherogram (**Figure 44**) shows crossing of peaks so if the variation is genuine or not is yet to be established. As illustrated it could be a heterogenous variations.

For exon 21 results, two samples (21_5 and 21_8) failed, whereas six samples (21_1, 21_2, 21_3, 21_4, 21_6, and 21_7) show perfect match and sequence identity match. For sample 21_9 the electropherogram data was of poor quality. Variation was found in sample 21_12 but outside the coding region. For samples 21_10, 21_11, and 21_13 the coding region aligned perfectly, however, the PCR product was not of the expected size and the flanking regions matched elsewhere leading to extended span of the alignment, which needs further testing.

For exon 22 results, only one sample (22_8) failed to yield results, four samples (22_3, 22_4, 22_5, and 22_10) had 100 percent identity match, while a further four samples (22_1, 22_2, 22_7, and 22_13) had variation outside of the coding region. For sample 22_6, a duplication (c.3848dupG) is found, which if genuine could lead to a frame shift at amino acid position 1282. However, there is no corresponding peak at that position on electropherogram results. For sample 22_9, the electropherogram had missing peaks in part affecting the sequence output and hence the alignment results. For sample 22_11, a substitution (c.3908C>T) is found, and if genuine it would lead to an amino acid substitution p.Pro1302Leu, again the electropherogram results shows possible heterogeneity in the sequence and it might not be a true substitution as both alleles could be present. Another variation is found outside of the coding region. For sample 22_12, two insertions are found, one outside of the coding sequence and the other c.3862insT which, if genuine, leads to a frame shift at amino acid position 1285. The electropherogram results show small peaks within the larger peaks.

For exon 23 results, four samples (23_1, 23_2, 23_8, and 23_10) show perfect alignment, while for samples 23_3, 23_4, 23_5, 23_6, 23_7, 23_9, and 23_11, the coding sequence aligned perfectly but there were variations outside the coding region

and more towards the primer sequence end. For sample 23_12 there was an insertion c.3951ins A causing a frame shift that affects the amino acid sequence and prior to that (of the translation end) there is a base substitution c.3953C>T leading to an amino acid change p.Pro1317Leu from a proline to a leucine. On the electropherogram, however, the corresponding peaks are very small and calling could be based on heterogeneity rather than genuine insertions. Last, for sample 23_13, a deletion c.3962 del A, affecting the amino acid sequence at position 1321.

For exon 24 results, six samples (24_3, 24_4, 24_6, 24_10, 24_11, and 24_12) show perfect alignment. While for samples 24_1, 24_2, 24_9 variation was outside of coding region. For the remaining samples (24_2, 24_5, 24_7, 24_8, and 24_13) the electropherogram data was of poor quality.

For exon 25 results, sample 25_1 and 25_2 had variation outside of the coding region. All other samples failed to generate results at PCR stage although the same conditions were used as of the previous two samples but were run on different days.

For exon 26 results, ten samples (26_1, 26_2, 26_3, 26_4, 26_5, 26_6, 26_8, 26_9, 26_12, and 26_13) had a complete alignment identity score. Samples 26_7 and 26_10 have no variation within the coding region and finally sample 26_11 had poor quality electropherogram results.

For exon 27 results, three samples (27_1, 27_2, and 27_3) have complete sequence identity score while for five other samples (27_4, 27_6, 27_8, 27_12, and 27_13) variation was outside of coding region. The electropherogram results for four samples (27_5, 27_7, 27_10, and 27_11). Last, for sample 27_9 there is a duplication c.4525dupA, which causes a frame shift at amino acid position 1508, however, looking at the electropherogram there is no corresponding peak to that duplication and might be a wrong base calling.

For exon 28 results, sample 28_9 failed, while eight samples (28_2, 28_4, 28_5, 28_8, 28_10, 28_11, 28_12, 28_13) show complete identity. For samples 28_1, 28_3, 28_6, and 28_7 have no variations within the coding sequence.

For exon 29 results, seven samples (29_1, 29_3, 29_5, 29_7, 29_10, 29_11, and 29_13) show 100 percent identity match and no variation was found in the coding region for samples 29_4, 29_6, 29_8, and 29_12. The quality of the electropherogram result was poor for sample 29_9. Finally, one base substitution is found for sample 29_1 (silent mutation c.4812C>T), which has no effect on the amino acid sequence.

For exon 30, four samples (30_1, 30_2, 30_5, and 30_8) failed to yield results, while seven other samples (30_3, 30_4, 30_6, 30_9, 30_10, 30_11, and 30_13) had a 100 percent identity match. Only two samples (30_7 and 30_12) had variations but outside of coding region.

A summary table of the results has been constructed **Table 31** below, showing the variations found within the coding region.

Table 31 lists a summary of the results of variations identified within the coding region

Exon	Sample(s)	Variation(s)
Exon 20	9	**c.3766T>A (p.Ser1255Thr)**
Exon 22	6	**c.3848dupG**
	11	**c.3908C>T (p.Pro1308Leu)**
	12	**c.3862insT**
Exon 23	12	**c.3951insA**
	12	**c.3953C>T (p.Pro1317Leu)**
	13	**c.3962delA**
Exon 27	9	**c.4525dupA**
Exon 29	1	**c.4812C>T**

4.4 Discussion

Only 9 variations were identified, with one found in a normal individual. All samples with variations will still require reruns to confirm if any human or base call errors occurred, while the majority of the other samples will need to be rerun, and in case of the variation within the coding region a further confirmatory testing is required of parents to establish if the variation is causative of the phenotype or it is an inherited polymorphism. Although mutations in CREBBP gene are usually *de novo*, it is possible for the parent to have the same mutation and be mosaic, which would not be visible as a phenotype (Roelfsema and Peters 2007). Reruns were not carried out due to shortages in reagents and delays in delivering new ones, but they are essential to establish if the variation is causative. A conclusion on the mutation detection rate can not be reached given the experimental design and runs caused by the limited supply of reagents.

As mentioned earlier some of the electropherogram results of the samples did not show distinct peaks that correspond to the variation and hence a rerun is vital to confirm the presence of the variation in question. In addition sequencing should be carried out for the reverse complement to make sure that the variation is genuine and not due to slippage. After the variation is confirmed the next step would be to test the parents to see if the variation is pathogenic.

One limitation of the analysis in question is the lack of matched control samples that would identify normal variations. One of the samples used was of a parent of an RTS patient and a variation **c.4812C>T** at exon 29 was found, but not in the affected child, that has no effect of the amino acid sequence and hence can be assumed to be benign. However, having more controls could offer better results for comparison and to identify problems in the experimental process.

Failed results might be due to many reasons such as problems with the primer set, quality of DNA sample and possible contaminants, as well as a potential deletion of the

exon in question as the primer binding site would be lost and hence no product would be produced. This would be suspected if one sample failed, as it is not very common to have the same exon deletions in unrelated RTS cases.

For the case of exon 25, sample failure is suspected to be due to primer expiry, primer dimer or a missing reagent in the master mix.

Focus in this analysis is on variations within coding regions at the first instance and within the exons covering the HAT domain in particular, which has been associated with main function of the protein. This does not mean that variations outside of the coding region will not be pathogenic as it is possible but are less likely. These could disrupt splicing and hence result in an incomplete transcript than required for proper protein function.

The variations identified here will require further follow up to determine pathogenity. Reruns are important especially in reverse order to determine that the mutation is genuine. Also testing parents or other unaffected sibs would confirm if the mutation is causative or just a polymorphism. All of the identified variations are novel and were not found in the literature however some might be technical artefacts and might be caused by incorrect base calling.

Pathogenic variations of other CREBBP exons have been reported and mutations may lay outside of the tested exons in these samples. In addition, it is also possible that the variation that is behind the phenotype on the epigenetic level and so would not be detected by direct sequencing. Another possibility is that the causative variation is in another gene such as EP300.

The research in this chapter has implications for the appropriate clinical genetics work flow in Kuwait to ensure that possible RTS mutations in CREBBP are investigated fully. Given the rarity of the disorder and the number of reoccurrences, a suggested work-flow would be one that is efficient and of low cost, chromosomal or genomic search of

possible deletions duplications that would identify abnormal regions that might account for the disorder. In the Kuwait Medical Genetic Center (KMGC), this is a routine check via karyotyping and using FISH probe and specifically for chromosome 16 where a deletion would be first suspected.

If no aberration is detected then the following steps nowadays is to go for genomic microarray technology or even next generation sequencing, which is not yet available in the KMGC and will require a long period to set up. However, there are options of sending samples abroad for analysis, but these usually assume some level of understanding of output data, as several candidate regions would be identified, unless the laboratory in contact has specific expertise in RTS sample, and these will require further research. Testing through direct sequencing for RTS gene mutations is important, even if as a confirmatory step, since two genes are already known and the primers can last for years.

Another local option would be to use RTS specific MLPA kits. As mentioned in chapter 4 earlier, MLPA is a more robust and easy to handle method to handle than sequencing and would test different exons at the same time. However, due to kit shelf life, samples need to be collected prior to ordering the kit, which might need a waiting period due to the rarity of the disorder.

When no exon deletion duplication is detected then direct sequencing would be another option and it would still be required for confirmatory purposes of MLPA results. Also the primers shelf live, assuming proper storage, can extend for years.

The sequencing study presented here is limited by technical restrictions. It could only be performed with the reagents supplied, and therefore repeat runs could not be made. The experiment requires more controls to have been added and testing of parents would have been carried out to confirm variations identified.

It is of great comfort for families with affected individuals to identify the causative genetic factor that might aid in better family planning and event possible intervention, with current attempts to target the CREBBP function through therapeutic drugs. In the case of RTS, occurrence is usually de novo but here two families have multiple affected children, which might suggest another mode of inheritance than previously reported.

Identifying the genetic cause of a rare disorder might be expensive and time consuming, however, it will help in further understanding of biological pathways and gene function that might aid in other related disorders or even for normal function. For instance, CREBBP gene has been associated with memory function and hence might be useful for other intellectual disorders and shed more knowledge on this brain process. Hence the value of researching rare disorders that can identify such novel genes and enhance our understanding of their cause.

KMGC is currently writing proposals to further investigate these cases. More details is provided in chapter 7.

Chapter 5 - MLPA Results

5.1 Background

Clinical genetic laboratories have traditionally used light microscopy to detect chromosomal aberrations in patents suspected of genetic disorders. However, recent improvements in technology, especially the development of array-based tools such as comparative genome hybridisation (array CGH), has led to the realisation that many genome deletions and duplications which cannot be seen by standard microscopy are important in human genetic disorders. Thus, for idiopathic intellectual disability and many other neurodevelopmental conditions many laboratories have made the shift from classical karyotyping and fluorescence *in situ* hybridization (FISH) analysis to array CGH.

An alternative or complementary screening tool to array CGH is the Multiplex Ligation-dependent Probe Amplification (MLPA) method. Studies have reported an increase in the detection rate of micro deletion/duplication using MLPA and have recommended it as a first step robust and reliable screening method and a cost-effective alternative to more costly methods such as array CGH or multi-probe Flurorescent In Situ Hybridisation (FISH) techniques, (Koolen, Nillesen et al. 2004; Rooms, Reyniers et al. 2004), (Northrop, Ren et al. 2005) (Palomares, Delicado et al. 2006), (Monfort, Orellana et al. 2006), (Ahn, Ogilvie et al. 2007), (Stegmann, Jonker et al. 2008), (Stuppia, Antonucci et al. 2012). A wide range of commercial MLPA kits is available that are disorder specific or region specific, and it is also possible to self-design MLPA assays.

MLPA is a PCR based technique used to identify copy number changes of up to 50 DNA sequence probes in one single multiplex reaction (Palomares, Delicado et al. 2006). The probes result in unique amplified PCR fragments that range in size between 64 to 500bp, which are quantified by capillary electrophoresis. Each MLPA probe comprises of two oligonucleotide fragments that needs to bind to adjacent target DNA sequence

for ligation to occur, this is then followed by amplification of the ligated single fragment probe. MLPA is different from standard multiplex PCR as only a single PCR primer pair is used for all of the probes.

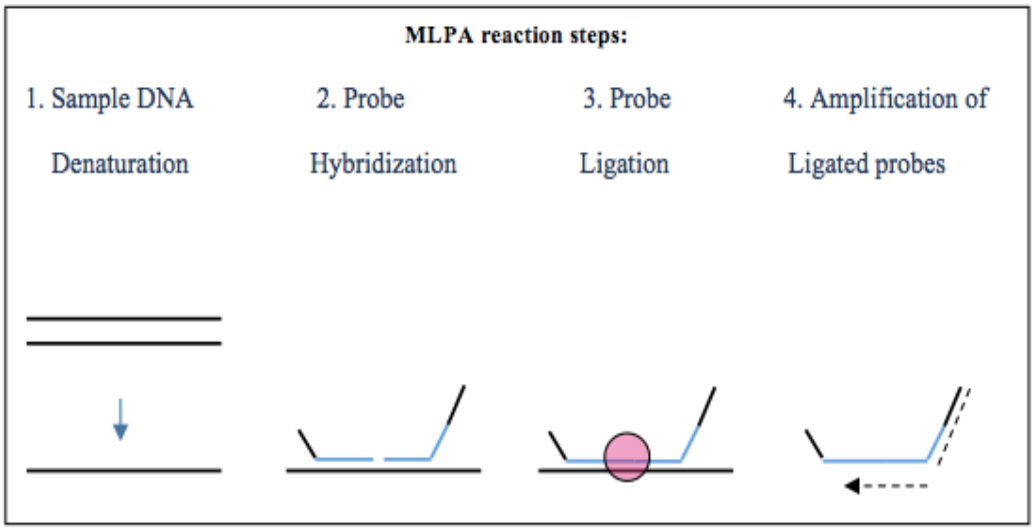


Figure 45 The steps involved in an MLPA reaction

Another important distinguishing feature of MLPA compared to standard multiplex PCR is that amplification is performed on ligated probes that have been hybridised to the sample DNA and not the amplification of the sample DNA itself, as illustrated in step 4 of **Figure 45**. Each probe is designed to generate a PCR amplicon that has a unique length in order to separate and quantify up to 60 different PCR amplicons (including quality control fragments) by capillary electrophoresis.

MLPA is a relative technique and only detects relative differences. Hence, the amount of sample DNA present is not important and gives no indication about copy number and it is rather the comparison with a reference DNA samples in the same run that generates the results of copy number status. It is important to select the reference samples (a normal control sample with no deletions or a duplications in the tested probes) carefully to achieve better results (Stuppia, Antonucci et al. 2012).

5.1.1 Quality control

Each one of the MLPA probemix kit contains control fragments to detect errors in the MLPA reaction.

5.1.1.1 Internal quality control

Internal quality controls are used to check if the reaction was successful by indicators such as DNA quantity, denaturation error detected and check for sample swapping. The list of internal quality controls and their purpose is given in **Table 32** below

Table 32 Internal quality control fragments, the size and purpose of each

Name	Length (nt)	Purpose
92 nt control	92	Ligation-dependent probe that is used as a benchmark to compare other control fragments to
Q-fragments	64, 70, 76, 82	Used as an indicator if the amount of DNA used was too low. If signals are higher than 1/3 of the 92nt control fragment then the DNA quantity will be regarded as insufficient
D-fragments	88, 96	These fragments detect denaturation problems. If signals are lower than 40% of the 92nt control fragments indicate a denaturation problem.
X & Y fragments	100, 105	Detects sample swapping by gender confirmation

Q-fragments

These consist of oligonucleotides that contain both MLPA PCR primers sequences in a single molecule; hence they do not need hybridisation or ligation to occur in order to be amplified during the PCR process. These fragments are present at small quantities

and are outcompeted by the amplicons of the MLPA probes when enough DNA is used. Low peaks of these Q-fragments is therefore an indication that enough DNA was used and ligation was carried out successfully. On the other hand, if the peaks are higher than a third of the height of the 92nt control fragment this indicates that the sample DNA amount was low or that the ligation failed.

D-fragments

These are synthetic probes that hybridise to a sequence within CG rich region that require high temperatures to denature. Hence, they serve as a measure of the success of the denaturation process. When the peaks of the D-fragments (88 and 96nt) are lower than 40% of the 92nt control fragment, this indicates that the denaturation of the sample DNA was incomplete. This could be due to salt contamination, and will result in unreliable results for probes that hybridise to sequences with close proximity (<5 kb) to CpG islands. These are usually present near transcription start points and hence, would mainly affect probes for the first exon.

X & Y fragments

X and Y fragments are sized 100 nt and 105 nt respectively and detect AMOT (Xq23) and UTY (Yq11.221) genes. These control fragments are used for sex determination and in some cases to detect sample swapping.

5.1.1.2 Quality control in the design of the experiment

A negative control sample (with no DNA) is added to reveal possible contamination (with DNA and RNA fragments) of dH_2O, TE, MLPA reagents, electrophoresis reagents or capillaries.

Multiple reference DNA samples could be used to measure reproducibility. At least 3 copies of the reference samples per MLPA run are recommended and should be distributed randomly over the sample plate. In addition, it is recommended to add one

additional reference sample for each seven DNA samples. In the case of this study three different reference samples were used but not distributed as recommended.

5.1.2 Using MLPA to detect the genetic cause of mental retardation

A number of MLPA kits are available from MRC-Holland for the detection of copy number gain or loss in gene regions attributed to mental retardation. Due to the fact that genes involved in the cause of mental retardation lie in over 100 genetic regions, no one kit can include all the regions.

For non-syndromic intellectual disability a workflow using the available kits is suggested by MRC-Holland, as illustrated in **Figure 46** below.

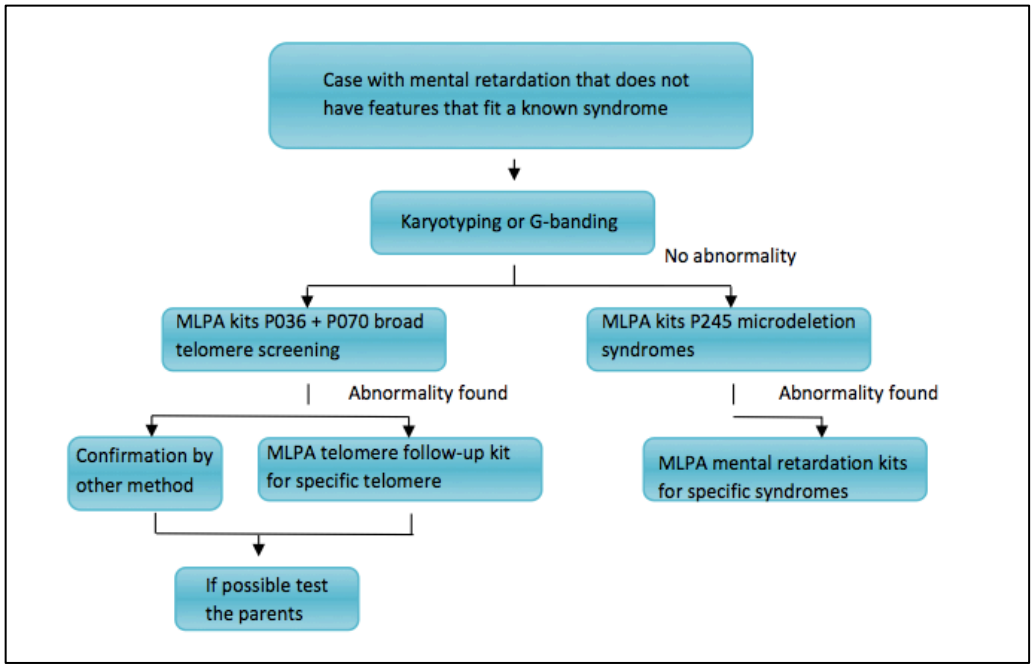


Figure 46 A workflow to screen causes of idiopathic intellectual disability, from MLPA Holland

Therefore, as recommended MLPA kits P245-A2, P036-E1 and P070-B2 were ordered for this study. Details of their probe content can be found in chapter 3.

5.2 Material and methods

Laboratory work was carried out using the materials and methods listed in chapter 3.

One DNA sample was used from each family. The choice of individual was based on the DNA quantity and having a concentration above the minimal required working concentration of 25ng/ul. Concentrations below this would require further concentration of the sample through speed vacuum and would affect the results. Based on these criteria, no samples were used from families 12, 14, and F. DNA from families 10, A, and H was not included due to shortage of sample.

In total 13 samples from affected members of idiopathic ID families were used and 7 samples from affected RTS families, in addition to three control samples with no family history of intellectual disability.

5.3 Data analysis

5.3.1 Evaluation of raw data

Before proceeding to probe ratio calculation the quality of the MLPA run was first evaluated as recommended by the manufacturer's by inspecting the electropherogram (MLPA peak pattern of each sample). This is done to remove any data that may affect or introduces artefacts to the analysis. After which, the run could be repeated. Only data that passed the raw data checklist and size-called data would be suitable for MLPA data analysis using the coffalyser.NET software.

Raw data was viewed using Beckman GenomeLab software and compared to the below checklist in **Table 33** below.

Table 33 Raw data checklist to evaluate results quality

No	Checklist question
1	Is there a high baseline?
2	Is there a high background signals and/or shoulder peaks?
3	Do the peaks of the size standard fragments differ in height?
4	Are all peaks very low?
5	Is there too much signal sloping?
6	Are spiky peaks present?
7	Are truncated peaks or very high peaks present?
8	Is there any broadening of peaks?
9	Is there spectral pull up/pull down patterns in colours?
10	Are there any irregular current patterns or decreasing current?
11	Are one or more peaks outside of the recommended signal range (for Beckman CEQ 3,000 rfu or 3x baseline height to 170,000 rfu)

This was then followed by the visualization of size-called data. First the internal quality control fragments were checked (64-105bp). Peak pattern was evaluated using a flow chart available from MRC-Holland (attached to the appendix). This was then followed by Dosage quotient (copy number) calculation using the Coffalyser.Net software.

5.3.2 Optimising MLPA peak patterns

An optimal peak pattern is one that has a low baseline (as straight as possible), minimal signal intensity that is 3 x the baseline intensity and 40% of the maximum intensity, an acceptable signal sloping is one in which the average signal intensity of the longest

fragments should be at least 1/3 of that of the shorter fragments, and peaks should not have shoulder peaks.

Signal intensity is affected by concentration of the MLPA products in the injection mixture, injection voltage and time. The following are general rules:

- Increased concentration of MLPA PCR product leads to increased probe signal intensity and vice versa, (except when the MLPA PCR product is more than 10% of the total volume as this causes an increase in salt content which will reduce signal intensity).
- Small decrease in the injection voltage or time leads to increased DNA uptake and increased intensity and vice versa. The effect of injection voltage is higher than that of injection time on signal intensity. Hence, injection time can be adjusted to improve signal intensity. A recommended injection time is between 15- 40 seconds.

Shoulder peak (+/- 1 nt of the main probe signal) is influenced by polymer (gel) type, run voltage and time. By leaving the PCR tubes with the MLPA product in the dark for an extra hour after PCR reaction, a +1 nt can be avoided as an extra adenosine nt will be added to all amplicons by the non-proofreading polymerase found in the MLPA mix.

A thick gel type has a low resolution, which is suitable for fragment analysis and peak quantification opposed to a high resolution gel that may cause shoulder peaks to be regarded as a separated peak, reducing the detected probe peak area. Alternatively reducing the run voltage decreases resolution and this can be balanced by increasing run time.

5.3.3 Coffalyser.NET

MRC-Holland, the suppliers of the MLPA kits, designed the Coffalyser.NET software for data analysis of the MLPA results. The analysis of each experiment comprises of two main stages: 1) raw data analysis, and 2) comparative analysis.

The raw data analysis process includes:

i) Baseline correction: non-specific fluorescence or background auto fluorescence is subtracted from the probe product fluorescence in order to calculate the relative fluorescence that is caused by the incorporation of the fluorophore.

ii) Peak detection- Peak detection algorithm is carried out in two steps. First step is by comparison of intensities of fluorescent units to a set of arbitrary thresholds and shape recognition models. Second, filtering of the generated list of peaks is applied by relative comparison.

iii) Peak size calling- In this process the detected peaks of the MLPA sample channels are compared against the selected size markers.

iv) Peak identification- This involves algorithm for peak to probe linkage using a window-based binning approach, in which all peaks within a given size window across different samples are considered to be the same peak and are matched to a specific peak size, e.g. a range of 6 nt of expected fragment size.

v) Raw data quality control- these include baseline height, peak signal intensity, signal to size drop, incorporated percentage of prime to expected standards specific for each capillary system. Other control checks involve the quality control fragments included within the kit. Also quality control scores are calculated e.g. fragment run separation score (FRSS) and fragment MLPA reaction score (FMRS).

The software will indicate if there was a problem detected in DNA quality or denaturation steps and samples that failed the checks are then excluded from the next process.

Following raw data analysis, comparative analysis is carried out on samples that passed the raw data checks. This starts out with a normalization step, which is performed to bring unknown and reference samples to a common comparative scale, this commonality is achieved by the division of multiple data sets by a common variable or normalization constant to cancel out the effect of that variable on the data. The algorithm in the software uses the reference probes in order to create a common variable. Then the data for each test probe of each sample is compared to each available reference sample producing as many dosage quotients as there are reference samples.

5.3.4 Normalisation

For data analysis, two normalisation steps are carried out. The first, by comparison of peaks of probes detecting genes of interest to the reference probes within the sample (intra-sample normalisation). The second, comparison of peak pattern of sample of interest to the reference sample DNA included in the same experiment (inter-sample normalisation). Both reference probes and the reference sample are assumed to have normal copy numbers at the specified locations. A reference sample should be from a healthy individual with no family history of the tested disorder. Abnormal signal probes will indicate a possible deletion or duplication within the sequence of interest. Therefore, careful reference sample selection is vital as it affects the MLPA results.

5.4 Results

Raw data analysis was carried out using Beckman GenomeLab software while comparative analysis was carried out using Coffalyzer version 1.0.0.43.

5.4.1 Beckman GenomeLab software results

The results were viewed using the Beckman GenomeLab software and only results that passed the raw data checks were copied over for analysis with Coffalyser.NET.

For kit P245-A2, 19 out of the 23 samples passed the initial results, for kit P036-E1, 20 out of 23 samples passed, while for kit P070-B2, 18 out of 23 samples passed.

All passed samples had some spectral pull as illustrated in **Figure 47** by the separated red and blue peaks generated for each probe signal. These may affect results slightly when it comes to peak calling and binning.

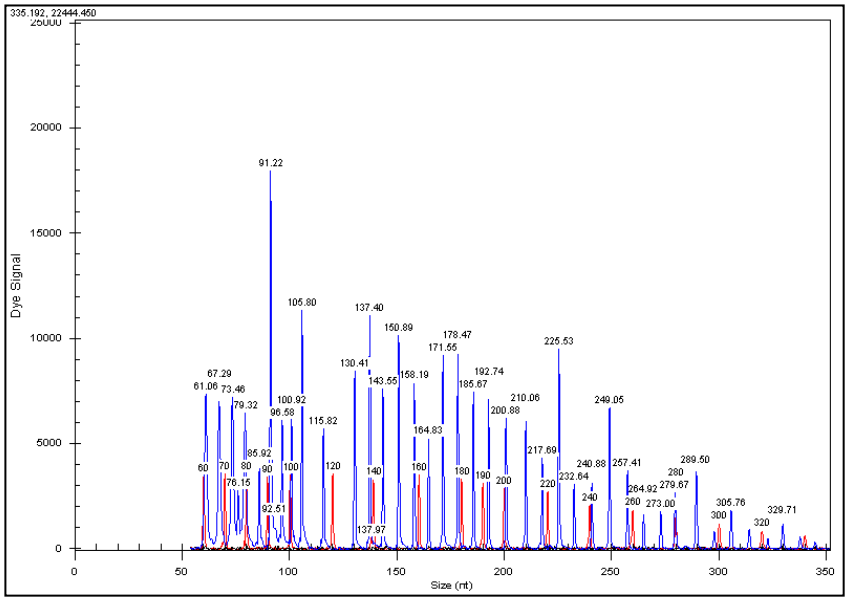


Figure 47 A snapshot of the results generated for the detected peaks of sample cc8 Little spectral pull up is showing (red to blue)

5.4.2 Coffalyse.NET results

5.4.2.1 Quality control results

Results table were generated with all the signal data as well as quality control results that include DNA quantity, DNA denaturation and detection of X- and Y- fragments. The following tables 31, 32, and 33 are constructed to include the quality control results as well as the sample number, gender and number of probes amplified, for each kit separately.

Table 34 Summary of the quality control results for kit P245-A2, including the sample number, gender and number of probes amplified, DNA quality, DNA denaturation (DD), and detection of X- and Y-fragments.

Sample	Gender	Probes detected	DNA	DD	X-Frag	Y-Frag
-Ve	NA	0 / 50	Too low	Not OK?	Not detected	Not detected
1.1	M	15 / 50	Too low	Not OK?	Detected	High
2.1	M	4 / 50	Too low	Not OK?	Detected	Not detected
4.1	M	44 / 50	OK	Not OK?	Detected	Not detected
5.1	F	35 / 50	OK	Not OK?	Detected	Not detected
6.1	F	44 / 50	OK	Not OK?	Detected	Not detected
7.1	F	20 / 50	OK	Not OK?	Detected	Not detected
9.1	M	19 / 50	OK	Not OK?	Detected	Detected
11.1	F	44 / 50	OK	Not OK?	Detected	Not detected
15.1	F	36 / 50	OK	Not OK?	Detected	Not detected
16.1	M	36 / 50	OK	Not OK?	Detected	Detected
B.1	F	36 / 50	OK	Not OK?	Detected	Not detected
C.1	F	42 / 50	OK	Not OK?	Detected	Not detected
D.1	M	17 / 50	OK	Not OK?	Detected	Detected
G.1	M	39 / 50	OK	Not OK?	High	High
I.1	M	44 / 50	OK	Not OK?	Detected	Not detected
CC1	M	42 / 50	OK	Not OK?	Detected	Detected
CC3	M	40 / 50	OK	Not OK?	Detected	Detected
CC8	F	44 / 50	OK	Not OK?	Detected	Not detected

Table 35 Summary of the quality control results for kit P036-E1, including the sample number, gender and number of probes amplified, DNA quality, DNA denaturation (DD), and detection of X- and Y-fragments.

Sample	Gender	Probes detected	DNA	DD	X-Frag	Y-Frag
-Ve	NA	0 / 47	Too low	Not OK?	Not detected	Not detected
1.1	M	26 / 47	OK	OK	Detected	Detected
2.1	M	47 / 47	OK	OK	Detected	Detected
3.1	M	47 / 47	OK	OK	Detected	Detected
4.1	M	46 / 47	OK	OK	Detected	Not detected
5.1	F	46 / 47	OK	OK	Detected	Not detected
6.1	F	46 / 47	OK	OK	Detected	Not detected
7.1	F	46 / 47	OK	OK	Detected	Not detected
9.1	M	29 / 47	OK	OK	Detected	Detected
11.1	F	46 / 47	OK	OK	Detected	Not detected
13.1	F	46 / 47	OK	OK	Detected	Not detected
15.1	F	46 / 47	OK	OK	Detected	Not detected
B.1	F	46 / 47	OK	OK	Detected	Not detected
C.1	F	46 / 47	OK	OK	Detected	Not detected
D.1	M	45 / 47	OK	OK	Detected	Detected
G.1	M	47 / 47	OK	OK	Detected	Detected
I.1	F	46 / 47	OK	OK	Detected	Not detected
J.1	F	44 / 47	OK	OK	Detected	Not detected
CC1	M	47 / 47	OK	OK	Detected	Detected
CC3	M	46 / 47	OK	OK	Detected	Detected
CC8	F	46 / 47	OK	OK	Detected	Not detected

Table 36 Summary of the quality control results for kit P070-B2, including the sample number, gender and number of probes amplified, DNA quality, DNA denaturation (DD), and detection of X- and Y-fragments.

Sample	Gender	Probes detected	DNA	DD	X-Frag	Y-Frag
-Ve	NA	0 / 47	OK	Not OK?	Not detected	Not detected
1.1	M	14 / 47	OK	OK	Low	Detected
2.1	M	46 / 47	OK	OK	Detected	Detected
3.1	M	46 / 47	OK	OK	Detected	Detected
4.1	M	45 / 47	OK	OK	Detected	Not detected
5.1	F	36 / 47	OK	OK	Detected	Not detected
7.1	F	45 / 47	OK	OK	Detected	Not detected
9.1	M	46 / 47	OK	OK	Detected	Detected
11.1	F	45 / 47	OK	OK	Detected	Not detected
13.1	F	45 / 47	OK	OK	Detected	Not detected
15.1	F	45 / 47	OK	OK	Detected	Not detected
16.1	M	41 / 47	Too low	Not OK?	Detected	Detected
B.1	F	45 / 47	OK	Not OK?	Detected	Not detected
C.1	F	45 / 47	OK	OK	Detected	Not detected
D.1	M	46 / 47	OK	OK	Detected	Detected
G.1	M	24 / 47	OK	OK	Detected	Detected
J.1	F	38 / 47	OK	OK	Detected	Not detected
CC1	M	44 / 47	OK	OK	Detected	Detected
CC8	F	42 / 47	OK	OK	Detected	Not detected

For kit P245-A2 all the samples had a denaturation problem (as indicated by the DD column in table 31). Two samples had a problem with DNA quantity and one male sample had errors in Y-fragment detection.

For all samples run with kit P036-E1, as summarized in table 32 the DNA quantity and denaturation was “OK” and the X- and Y- fragments were detected as per gender except for sample 4.1, where the Y-fragment was not detected although the sample was from a male case. The number of probes detected compared to the total (47

probes) ranged from 26 probes for sample 1.1 to all 47 probes for samples 2.1, 3.1, G.1 and CC1.

For all samples run with kit P070-B2, as summarized in table 33 the DNA quantity and denaturation was “OK” except for 16.1 where the DNA quantity was too low and denaturation was not “OK”, while for sample B.1 only denaturation was flagged as “Not OK”. All the X- and Y- fragments were detected as per gender except, again, for sample 4.1, where the Y-fragment was not detected although the sample was from a male case. The number of probes detected ranged from 14 probes for sample 1.1 to all 46 probes for sample D.1.

5.4.2.2 Probe results

Dosage quotient (DQ) or ratio is a measure for the ratio in which the target sequence is present in the sample DNA compare to the reference DNA. If *a* and *b* are the signals from two amplicons in the sample DNA and *A* and *B* are the corresponding amplicon in the reference then;

$$DQ = (a/b)/(A/B)$$

DQ is the main value to look at for data interpretation and copy number status can be inferred as listed in **Table 37** below. The reliability of the results is indicated by an overall standard variation per probe in the reference sample to be less than 10% (<0.1).

The Coffalyser.NET software has different results options, out of which a pdf summary of the results per sample can be extracted with three panels shown in **Figure 48**. On the top left is a graphical representation of the signal level against probes lengths. On the bottom left is a plot of the probes aligned by their mapview location against the Dosage Quotient, which is labeled ratio in the bottom graph on. The word ratio will be used throughout the thesis to refer to the Dosage Quotient.

Table 37 Classification of copy number status by Dosage Quotient (DQ)

Copy Number status	Dosage Quotient
Normal	0.85< DQ < 1.15
Heterozygous duplication	1.35 < DQ < 1.55
Homozygous duplication	1.70 < DQ < 2.20
Heterozygous deletion	0.35< DQ < 0.65
Homozygous deletion	0
Equivocal copy number	All other values

On the right hand side, a table summary of each probe result is produced to include probe details such as gene-exon, chromosomal location, mapview, length, signal, the ratio value, standard deviation value and a comparison of the ratio to the sample population (which is a comparison ratio of the samples in a single experiment or run, labeled Sam in the table).

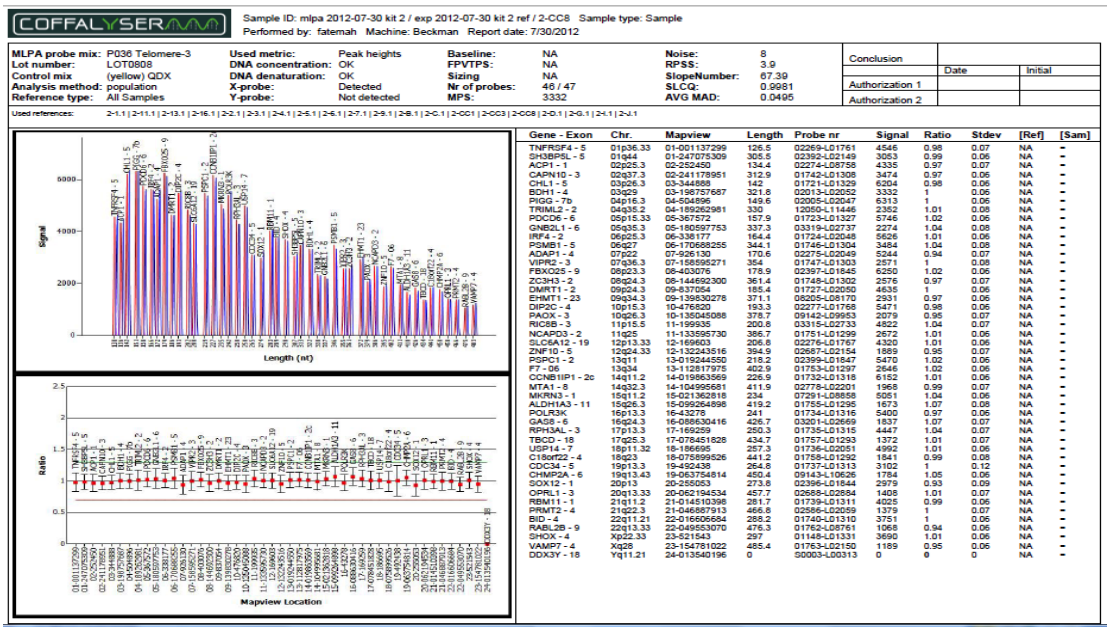


Figure 48 A snapshot of the pdf overview output results for sample CC8 run with kit P036-E1, including signal against probe length graph, ratio against mapview location graph and a summary table of each probe including ratio values and standard deviation.

Figure 49, next page, shows the graphical plot of the ratio against the mapview location for each probe for one case, for this example a sample from an control female case was used. As illustrated all probes have the ratio within the range of 0.7 to 1.3, except for the last probe, which has a ratio of 0, as expected since it is targeting a region located on the Y-chromosome and so would not amplify for a female sample. For each probe in each sample the ratio value was looked at first; for any probe with a ratio outside the expected range, the standard deviation was then checked, to ensure it was under 0.1 and if so the probe gene-exon and matching ratio noted in a summary table 35 and 36.

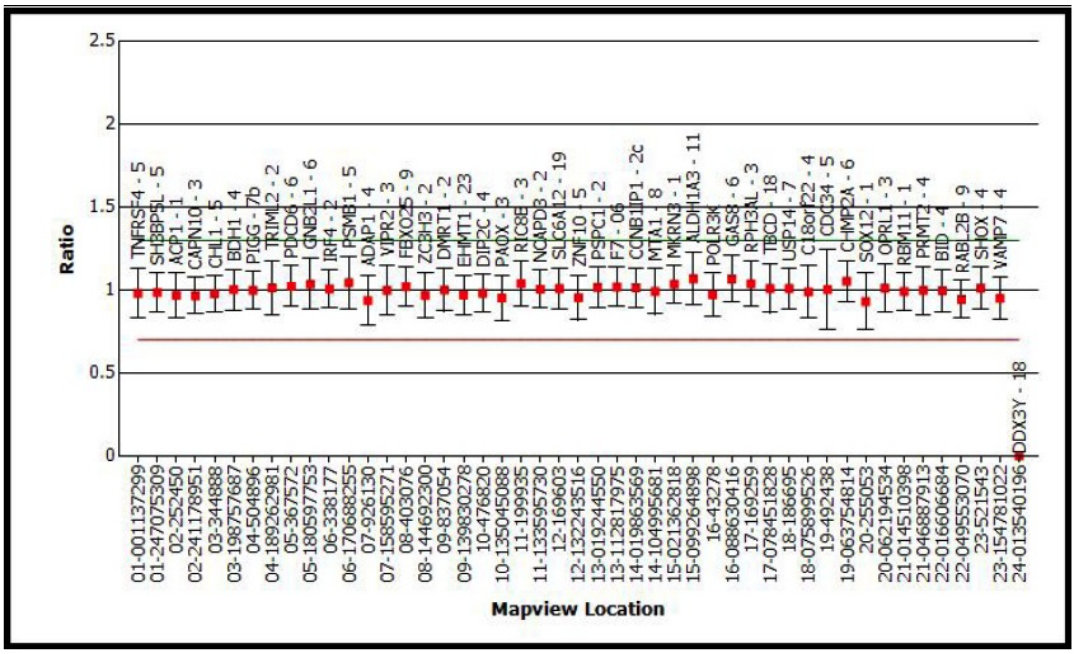


Figure 49 showing a normal plot of the ratio for each probe, arranged by Mapview location, for sample CC8 using kit P036-E1.

The two telomeric kits used P036-E1 and P070-B2 are meant to be used together, as confirmatory testing the either involved probes from the same gene but different exons or overlapped in the region. **Table 38** below shows the probes in each kit aligned side-by-side based on chromosomal position.

Looking at the ratio values for kit P245-A2 (results not shown) 13 out of the original 23 samples produced probe results at this stage, with six probes (TNFRSF4-5: 1p36.33, REL-9: 2p16.1, PAX6-6: 11p13, CLDN5-1: 22q11.21, DMD-35: Xp21.2, and MECP2-04: Xq28) showing a ratio value of 0 or above 1.3 across all samples including controls. Other probes also showed a ratio value of 0 across at least four samples. One probe SNAP29-1: 22q11.21 had a ratio value of 0 in only sample 5.1.

For kits P036-E1 and P070-B2, some samples failed the final analysis and some were dismissed for having a high number of probes with a 0 ratio value. For all kits, sample 4.1 (which is a male) did not have a ratio value higher than 0 for the DDX3Y-18:

Yq11.21 probe, which is located in the Y chromosome and should be amplified in male cases. This indicates a possible sample swap or a genuine deletion of that probe site.

Table 38 the probes used in each kit based on chromosomal position

	kit P036 E1 0910		kit P070 B2 0111	
chr. pos.	gene	exon	gene	exon
01p36.33	TNFRSF4	5	TNFRSF18	4b
01q44	SH3BP5L	5	SH3BP5L	3
02p25.3	ACP1	1	ACP1	5
02q37.3	CAPN10	3	ATG4B	7
03p26.3	CHL1	5	CHL1	3
03q29	BDH1	4	KIAA0226	22
04p16.3	PIGG	7b	PIGG	8
04q35.2	TRIML2	2	FRG1	1
05p15.33	PDCD6	6	CCDC127	3
05q35.3	GNB2L1	6	GNB2L1	2
06p25.3	IRF4	2	IRF4	3
06q27	PSMB1	5	TBP	2
07p22.3	ADAP1	4	UNC84A	5
07q36.3	VIPR2	3	VIPR2	2
08p23.3	FBXO25	9	FBXO25	8
08q24.3	ZC3H3	2	RECQL4	17
09p24.3	DMRT1	2	DOCK8	23
09q34.3	EHMT1	23	EHMT1	10
10p15.3	DIP2C	4	ZMYND11	2a
10q26.3	PAOX	3	ECHS1	8
11p15.5	RIC8B	3	BET1L	3b
11q25	NCAPD3	2	IGSF9B	20
12p13.33	SLC6A12	19	JARID1A	23
12q24.33	ZNF10	5	ZNF10	5
13q11	PSPC1	2	PSPC1	1
13q34	F7	6	CDC16	8
14q11.2	CCNB1IP1	4	PARP2	16
14q32.33	MTA1	8	MTA1	7
15q11.2	MKRN3	1	NDN	1
15q26.3	ALDH1A3	11	TM2D3	3
16p13.3	POLR3K		DECR2	9
16q24.3	GAS8	6	GAS8	11
17p13.3	RPH3AL	5	RPH3AL	2
17q25.3	TBCD	18	SECTM1	4
18p11.32	USP14	7	THOC1	21
18q23	C18orf22	4	CTDP1	9
19p13.3	CDC34	5	PPAP2C	7
19q13.43	CHMP2A	6	CHMP2A	3
20p13	SOX12	1	ZCCHC3	1
20q13.33	OPRL1	3	UCKL1	6
21q11.2	RBM11	1	HSPA13	2
21q22.3	PRMT2	4	S100B	2
22q11.21	BID	4	IL17RA	4
22q13.33	RABL2B	9	ARSA	1
Xp22.33PAR	SHOX	4	SHOX	5
Xq28PAR	VAMP7	4	VAMP7	8
Yq11.21	DDX3Y	18		

For kit P036-E1, ratio values of zero were found for two probes RABL2B-9: 22q13.33 and VAMP7-4: Xq28 in two RTS cases, D.1 and J.1 and form RABL2B-9: 22q13.33 in one control sample CC3.

For kit P070-B2, 8 probes showed variable copy number, with probe HSPA13-2: 21q11.2 having 0 ratio value across all samples. Four samples were dismissed and 5 samples failed leaving probe results for a total of 12 samples. For sample 3.1, the ratio value of the SHOX-5: Xp22.33 probe was less than normal at a value of 0.68, however, the corresponding standard variation value was more than 0.1 (values shown in the appendix). A similar pattern of results was seen for sample 11.1 where all 5 probes with ratio values less than 0.7 had a standard variation of over 0.1 (indicating the results might not be reliable). A ratio value of less than 0.7 is indicative of a heterozygous deletion.

For the control samples CC1 and CC8, probes ZCCHC3-1: 20p13 had a ratio value of 0 in both samples and probe IL17RA-4: 22q11.21 had a 0 ratio value for sample CC1. For CC3, a 0 ratio value was observed in probes S100B-02: 21q22.3 and SHOX-5: Xp22.33

Table 39 Results for kits P036-E1 and P070-B2, showing gene, exon number and ratio values (DQ) for samples with a ratio value outside of the normal range and with an standard deviation (SD) values <0.1.

Sample	Gender	MLPA P036-E1		MLPA P070-B2	
		Gene-exon: location	DQ	Gene-exon: location	DQ
3.1	M	**normal**		**SHOX- 5: Xp22.33**	**0.68**
11.1	F	**normal**		**CHMP2A-3 : 19q13.43**	**0.66**
				ZNF141-1 : 4p16.3	**0.67**
				S100B -02 : 21q22.3	**0.66**
				SHOX-5 : Xp22.33	**0.57**
				VAMP7-8 : Xq28	**0.57**

D.1	M	**RABL2B-9 : 22q13.33**	**0**	**normal**	
		VAMP7-4 : Xq28	**0**		
J.1	F	**RABL2B-9 : 22q13.33**	**0**	*dismissed*	
		VAMP7-4 : Xq28	**0**		
CC1	M	**normal**		**ZCCHC3-1 : 20p13**	**0**
				IL17RA-4 : 22q11.1	**0**
CC3	M	**RABL2B-9 : 22q13.33**	**0**	*failed*	
CC8	F	**normal**		**ZCCHC3-1 : 20p13**	**0**
				S100B -02 : 21q22.3	**0**
				SHOX-5 : Xp22.33	**0**

5.5 Discussion

Evaluation of the use of the MLPA kits selected in this research had limitations. The lack of confirmatory testing makes it difficult to establish if the deletion or duplication detected is valid or even causative. In this section, detection rate and the limitations of the current study will be discussed. The results will be compared to published data from the same kits, and finally recommendations for using MLPA kits will be made.

Denaturation was a problem mainly observed in results of kit P245-A2 and several probes had a zero value across samples, which is usually indicative of a denaturation problem or a homozygous deletion. It is more probable that the former is the case, in which denaturation failure would mostly affect sequences that are GC rich in content. This leads to failure of the DNA to denature properly, which will not allow successful hybridization, and consequential ligation and amplification. All of the probes correspond to known syndromes and for one patient to contain a multiple of syndromes is very rare. Therefore, the discussion will focus on the use of kits P036 and P070 from this point.

The abnormalities of the probes in these kits should be further confirmed by another method such as FISH, LR-PCR, Southern blotting or sequencing. Follow-up MLPA kits are also available that contain more probes per telomere for a closer examination of a specific region that would enable the confirmation of the aberration and to determine the size.

5.5.1 Detection rate

For the successful results generated (i.e. removing failed or dismissed data), an overall detection rate of around 17% (3 out of 18) was found, using kit P036, and around 33.3% (4 out of 12) using kit P070.

The detection rate of CNVs varies across different studies depending on sample size, selection criteria and kits used or detection methods. Detection rate must be defined as it could refer to specific types of aberrations or refer to an overall detection of all aberrations in a sample group, those that are clinically significant or to those that are detected by a specific kit or in a subtelomeric region.

In contrast lower detection rates were reported. Lam and colleagues screened a cohort of 20 unrelated patients, ranging from moderate to severe MR. First using MLPA for subtelomeric probes and then confirmed by FISH. For 20 patients tested 3 subtelomere deletions were found 1p36, 1q44 and 10q26, which yield a detection rate of 15%, which is higher than the reported 10%. The authors attribute this to the stringent inclusion criteria applied (Lam, Lam et al. 2006). While Christofolini and colleagues reported a detection rate of 7.6% of aberrations in subtelomeric regions, in their cohort of 132 patient, using MLPA P070 kit and qPCR (Christofolini, de Paula Ramos et al. 2010).

In addition, Mandal and colleagues reported an overall detection rate of 4.6% using kits P036-B and P070 (Mandal, Boggula et al. 2009) Other methods of whole genome

screening can lead to an increased detection rate of around 20%, while FISH based technique can detect 1.8% to 13.3% subtelomeric aberrations (Mandal, Boggula et al. 2009).

Moreover, Stegmann and colleagues reported an increase in detection rate of cryptic unbalanced chromosomal aberrations by 47% using MLPA as a detection technique with a detection rate of all types reaching a total of 12.2% including ID patients with or without multiple congenital abnormalities (Stegmann, Jonker et al. 2008).

Interestingly, Koolen and his team reported a detection rate of clinically significant aberration, using kit P036, to be varied based on severity of mental retardation of 6.3%, 5.1%, and 1.7% corresponding to mild, moderate, and severe mental retardation respectively (Koolen, Nillesen et al. 2004). Again, a study by Rooms and colleagues reported a detection rate of subtelomeric aberrations to be 5.3%, using two different kits than used in this study, with a cohort size of 75 affected individuals (Rooms, Reyniers et al. 2004). This is much lower than the detection rate reported in this study.

Furthermore, Jehee and colleagues using karyotyping reported the highest detection rate of aberrations, 30%, using a combination of MLPA kits (P036, P070, and P064), however, their cohort included patients with multiple congenital anomalies and mental retardation (Jehee, Takamori et al. 2011).

Recently, Pohovski and colleagues reported 14% detection rate in a study of 150 patients with unexplained developmental delay/intellectual disability. The authors recommended MLPA as a cost effective first tier test for large cohorts (Pohovski, Dumic et al. 2013).

The detection rates reported of these studies are summarized in **Table 40** below.

Table 40 Summary of detection rates reported by different studies, including cohort size, and the MLPA kits used in each.

Study	Cohort	Detection method and MLPA kits used	Detection rate
(Lam, Lam et al. 2006)	20 unrelated patients, ranging from moderate to severe MR.	First using MLPA for subtelomeric probes and then confirmed by FISH.	15%
(Christofolini, de Paula Ramos et al. 2010).	132	P070 kit and qPCR	Overall 7.6%
(Mandal, Boggula et al. 2009)	65	P036-B and P070	Overall 4.6%
(Stegmann, Jonker et al. 2008)	466	P036B and P070	12.2% (ID patients with or without multiple congenital abnormalities)
(Koolen, Nillesen et al. 2004)	210	P036	Varied based on severity of mental retardation of 6.3%, 5.1%, and 1.7% corresponding to mild, moderate, and severe mental retardation respectively
(Rooms, Reyniers et al. 2004).	75	two different kits than used in this study	Subtelomeric aberrations 5.3%,
(Jehee, Takamori et al. 2011).	261 (multiple congenital anomalies and mental retardation)	P036, P070, and P064	30%
(Pohovski, Dumic et al. 2013)	150 (unexplained developmental delay/intellectual disability)	P036, P070 and P245	14%

In comparison, the detection rate observed in the present project was much higher. This is mainly due to the following reasons: a) The selection criteria was specific in having families with a history of ID and therefore a high probability of a genetic factor, b) the pre-selection criteria which included test results of normal karyotype and no fragile X expansion mutation, meant that micro-aberrations where still highly probable and might have been missed through karyotyping as it is limited by a 1Mbp resolution. c) Population/family specific polymorphisms could lead to failure of the probe to hybridize and hence might result in a ratio indicative of an aberration as a single base difference affects the results d) The normalization method, reference sample and analysis software used could have given rise to false positives. SNPs/mutations within the sequence detected by a probe can result in false positive results as well as differences in DNA purity between sample and reference DNA sample. e) Human error and poor analysis technique, as each sample was tested once with each kit and being used for the first time, not all aberrations can be considered true or attributed to the phenotype, they could be done to human error or analysis error that needed a manual correction f) Most importantly, the detection rate here was based on a very small sample size where many samples even failed to give results and hence, the detection rate is inflated and not really representative of the population in question.

Usually, in other studies several of the detected regions of aberration would be detected in chromosome specific probes used by both kits since they are close in targeted region. This was not the case of any of the samples used, yet is does not mean that the aberration is not clinically significant or is a false result, but having the aberration shown with the other kit can be regarded as a confirmatory step. Ahn and colleagues, reported clinically significant aberrations that were the result of a single probe (Ahn, Ogilvie et al. 2007) .The regions detected by each kit on the same chromosome do not target the exact region and hence an aberration might not always

show in both kits. Other confirmatory testing that targets the exact region of aberration must be used before reaching a conclusion.

Also, finding out the detection rate of aberrations using these two kits should be established with a larger cohort and include samples of known deletions.

5.5.2 Comparing results

The underlying assumption is that the causative CNV is due to a homozygous deletion inherited from heterozygous parents, in which having consanguineous parents increases the risk.

There are some probe variations across the kits used in various studies since different versions of the kits were used as MRC-Holland improved some kits over the years. However, not all aberrations are causative of the disorder in question.

Looking at the results of the current study, two probes seem interesting in kit P036-E1 as they only have ratio values of 0 in two RTS cases. These are RABL2B-9: 22q13.33 and VAMP7-4: Xq28. The former having a 0 value in control sample CC3 as well. Mutation in the RABL2B gene has been reported in a girl with mental retardation by Zhu et al 2009 (only abstract was viewed as the paper is in Chinese). A duplication of the probe located on chromosome 22q containing RABL2B gene, using kit P036-B, which is a different kit than the one used here (Ahn, Ogilvie et al. 2007). Another study reported a deletion in the probe on 22q11 (Rooms, Reyniers et al. 2006), while Jehee and colleagues reported a deletion in 22q using the same kit to be causative in one patient as well the Xp and Xq probe deletion in another sample (Jehee, Takamori et al. 2011). In addition, Koolen reported a deletion in the 22qter probe in one patient with severe mental retardation out of a cohort of 210 cases (Koolen, Nillesen et al. 2004). Nevertheless, the same deletion was observed in the normal control in this study, which was not used as reference and data is not included here.

For these samples repeating the experiment would be important as well as checking if the sample result is found in parents and or siblings. Further testing with specific gene kits would be required as a homozygotic deletion of one exon might not be pathogenic if alternative splice forms exist. Also a polymorphism could have lead to the deletion, which would explain the same deletion found in the control.

For sample 4.1, a 0 ratio value for the Y-chromosome across all kits was found and may indicate a sample a swap, as the expected gender was male (data not shown).

For kit P070, three samples (3.1, 11.1 and CC8) have an aberration in the SHOX: Xp22.33 probe. SHOX gene has been associated with short stature and since it is found on the X chromosome this might explain the ratio value observed in the male sample. Sample 11.1 has several probes with a heterozgyous deletion or having a value that is very close to the cut off value and some of these probes maybe regarded as normal. An interesting homozygous deletion is that of ZCCHC3: 20p13 probe which was found in two of the control samples. This is probably due to a polymorphism or a denaturation error. Again these samples must be rerun and relatives tested to establish clinical significance of the results.

5.5.3 Limitations

A number of factors caused limitations in the identification of clinically significant aberrations in the current study. These include limited experimental design due to lack of reagents, small sample size, lack of resources to carry out other confirmatory testing, lack of a local technical support team, the relative and sensitive nature of MLPA technique.

The budget for this project was limited and so although access to more samples was possible, the sample size was limited to the availability of reagents. With regards to technical problems such as the observed spectral pull-up/pull-down in all cases can be

due to uncalibrated spectral or matrix as well as replaced or realigned optics or having used a new type of polymer. This can be addressed by making sure the proper dye set/filter set is used. Spectral calibration is required when optics are changed, a new polymer type or dyes are used. This required a local technical support and machine maintenance, which was lacking during the course of the project.

For the other subtelomeric kits probe HSPA13-2: 21q11.2 had 0 value across all samples. There could be a problem with detecting this specific probe including binning errors in which probes will be wrongly binned or missed during size calling and binning process. The lack of technical support in analysis made it difficult to interpret results and make any manual adjustments to the binning setting.

MLPA technique is very sensitive and is more sensitive to contaminants compared to simple monoplex PCR assays. Contaminants include salts, phenol, ethanol, heparin and Fe. Hence, samples compared should be extracted from the same tissue type, using the same extraction method, have similar concentration and have undergone similar storage and treatment conditions. This was hard to guarantee as some samples were extracted and stored by the Genetic Centre technicians without recording the extraction method.

In terms of sample quality, DNA samples can be concentrated using ethanol precipitation rather than concentration via evaporation or SpeedVac method that might result in very high EDTA concentration, which affects the ligation and PCR reaction. Therefore, when the concentration of the proband DNA sample was low the DNA sample of another sibling was used when possible. Initial concentration was variable amongst the samples, ranging from around 15 to around 1000 ng/µl. For the initial MLPA step the samples were diluted and all samples were brought to the same concentration, however, the salt content would be different as different amount of TE buffer was used to dilute the DNA. This has an effect on the MLPA results.

Moreover, some families refused to give blood when there was an existing sample in the DNA bank and although they consented to the study we had to use older samples.

It is worth noting that MLPA kits will not detect inversions or balanced translocations since this would not result in copy number change but might contribute to pathogenicity due to displacement of transcription sites that will lead to no gene product.

Further more testing was not repeated including, failed tests, nor samples from the parents were tested to determine de novo status, which is important to establish association to the disorder, as well as having another validation method to confirm detection of the variation.

5.5.4 Recommendations

In agreement with previously published work, MLPA was found to be a robust and easy to handle technique in terms of laboratory set up and work load, provided it is set up with the right starting material i.e. samples extracted with the same DNA extraction method, a positive control, and confirmed reference samples with no deletions/duplication. It was possible to detect a number of CNVs, however, further confirmatory testing is required before the CNV can be linked to intellectual disability.

It is very important to be clear about the detection rate and designing a screening setup for increasing the detection rate of clinically significant aberration, which could be the result of combined techniques. The use of multiple kits would increase the detection rate but care should be taken in selecting the kits per individual and using the appropriate kits to confirm or screen samples of patients.

Several published studies used either a sample of normal individuals or those affected with known aberrations or a combination of both to test the accuracy of MLPA and

identify common polymorphisms (Koolen, Nillesen et al. 2004; Rooms, Reyniers et al. 2004). This was not carried out due to delays in reagent arrivals and limited quantity available for the set budget. However, this should be the first step included in kit optimization. Additional kits are available such as kit P064 that detects some common microdeletion syndromes.

Since the focus of this part of the project was on kits that target subtelomeric aberrations, it is worth mentioning the de Vries checklist for patients with submicroscopic subtelomeric aberration (de Vries, White et al. 2001). This checklist scores items that include family history of mental retardation, prenatal onset growth delay, postnatal growth abnormalities, more than two facial dysmorphic features, and congenital abnormalities. The resulting score will later help in determining which samples are selected for this detection method. This was referred to in several of the reported studies and was a basis of their selection criteria.

Other methods exist for screening subtelomeric aberrations. However, MLPA could be a suitable first step in a screening process. However, this recommendation for MLPA use would require technical and analytical support, further screening on a larger sample group and a proper confirmatory testing set up, where parents can be tested as well. A cost-efficient approach is achieved by using FISH as a confirmatory method in which only the probes that match the MLPA subtelomeric region of gain or loss is tested. Once confirmed, the same probes can be tested with prenatal samples. As carried out by Lam and colleagues who used MLPA on amniotic fluid cells P036 kit was used as well as P019/020 kits as confirmatory kits and showed normal results and subsequently the couple had a normal child (Lam, Lam et al. 2006).

Advanced array-MLPA has also been used for rapid screening as those reported by Yan and colleagues (Yan, Xu et al. 2011), who developed an array-based MLPA to screen for common aneuploidies with 116 tag-probes that span across regions on chromosomes 13, 18, 21, X and Y and 8 control autosomal genes.

This will require a specific set up that is more advanced than the normal MLPA methods. This could allow a certain degree of customization when selecting probes and might be efficient at looking at several regions at once in detecting aneuploidies.

Identifying a common causative aberration would be very useful for families with a history of intellectual disability, where related couples could be counselled before marriage or conception and where existing parents want to opt for prenatal diagnosis and even pre-gestation diagnosis. Hence, having a proper set up that can enable better diagnosis is highly recommended for a national diagnostic genetic laboratory. In addition, the extended family relation and large family size means that informative results will impact more individuals.

Establishing if an aberration is causative must be approached with caution and enough scientific evidence that are supported by expert opinion. This will require the workflow that combines multiple techniques, e.g. starting with MLPA and then moving on to real time PCR, CGH array or FISH technique or even starting out with the CGH arrays and then using the related MLPA kit as a confirmatory test. Testing the parents and other relatives, if possible, to link the aberration with affected status would then follow.

In conclusion, MLPA is a well-established technique that once set up with the appropriate support system can serve as a rapid screening method to detect around 10% of cases with idiopathic intellectual disability. A proper experimental set up and analysis is essential for result interpretation and identifying clinically significant aberrations.

The MLPA kits were useful as a first step screening for aberrations following karyotyping. However, further studies are required on healthy individuals to establish if there are gains or losses that are population specific and also to follow up with sequencing to identify SNPs or benign mutations that might have contributed to the probe loss results. This population frequency of CNV has to be established for each kit,

as some probe duplications have been reported in normal individuals in other populations.

In addition it is essential to test parents for de novo status or to establish if the variation is a population specific polymorphism. Furthermore tests need to be repeated to confirm results including, failed samples, as well as having another validation method to confirm detection of the variation.

Also a standard protocol should be set up in a laboratory which is equipped with all supporting reagents and machine maintenance and technical support that would optimize result quality.

A better experimental design will be to optimize the screening aberrations using the kits on normal sample group before testing samples from affected individuals.

These kits were never used in Kuwait before and hence SNPs that have not been reported before and are population specific might cause results indicative of a deletion of a specific probe while the results could be due to a SNP that have affected the hybridization or ligation site of the probe. In addition, a deletion of a probe does not mean that it is pathogenic. It could be a population specific SNP or CNV. Each individual has several CNVs, across the genome, which does not contribute to a known syndrome. The MLPA kits are initially designed by MRC-Holland with the target sequence being a region that is outside of reported CNVs in the normal population based on the Database of Genomic variants, which is a repository for CNVs in normal individuals. However, the Kuwaiti population is not represented and also there is no guarantee that the normal individual does not have a low IQ score or that there is no family history of intellectual disability.

No conclusions should be reached about abnormalities detected by MLPA probes without validation by another technique. The MLPA probes were designed for research purposes and cannot be used for diagnostic purposes without validation and

comparison of published literature. For this purpose it is recommended that handling of data interpretation to be done by an expert and hence training on the analysis and significance of the aberration is as vital as the laboratory set up.

Currently there is no rapid screening program for idiopathic intellectual disability in the Kuwaiti Genetic Centre and hence MLPA could offer the cost effective solution for large cohort studies and population screening.

Chapter 6- Cytogenetic array analysis

6.1 Background

Intellectual disability (ID), historically known as mental retardation (MR), is a highly heterogeneous group of disorders, with many known environmental and genetic causes. In addition to ID, affected subjects can also have syndromal symptoms, including behavioural problems, dysmorphism, growth abnormalities, heart defects, epilepsy, and other medical problem such as deafness. Microscopy detects large-scale (>3Mb) chromosomal abnormalities, affecting many genes, in up to 10% of cases. Although many X-linked, chromosomal and genomic forms are now well characterised (e.g. Fragile X, Down syndrome and Williams syndrome), most cases display no visible cytogenetic abnormality and remain idiopathic. Heterogeneity has hampered genetic analysis, especially for the autosomal chromosomes, where most genetic changes associated with ID are expected to lie (Ropers 2008).

However recent developments in technology, particularly DNA genotyping arrays and array comparative genomic hybridisation (aCGH) have allowed the identification of many previously undetectable microdeletions and microduplications, collectively known as copy number variants (CNVs). These involve DNA deletion or duplication segments greater than 1kb. CNVs are now being detected in ID (Morrow 2010). These are mainly large *de novo* variants, i.e. new mutations, which are very rare in the normal population, and involve the loss or gain of one copy of a genome region (Webber et al., 2009), and these are mainly dominantly acting. However, there are also CNVs which give rise to ID through recessive mechanisms, i.e. through loss of function. Recessive phenotypes involving homozygous and compound-heterozygous deletions and mutations have been reported in neurexin 1 and CNTNAP2 gene in cases with severe ID (Zweier, de Jong et al. 2009).

This chapter will evaluate the use of Affymetrix cytogenetic arrays to test samples of ID and RTS from the Kuwaiti Genetics Centre, and try to identify CNVs and homozygous regions that may harbour autosomal recessive mutations in the families studied. Autosomal recessive mutation will be the focus through CNV analysis and looking at homozygosity mapping, assuming the underlying mechanism is homozygous or compound-heterozygous variation. Further confirmatory testing is carried out for one family using exome sequencing.

Assessing the use of these arrays for potential use in diagnosis in clinical genetics setting is another aspect as well as developing recommendations for experimental design and workflow.

6.1.1 Affymetrix Cytogenetic 2.7M arrays

The Affymetrix Cytogenetic 2.7M array was introduced to the market around 3 years ago, but has not previously been used for the detection of CNVs in ID in the Kuwaiti population.

The arrays are designed to screen for aberrations at a genomic level and more detail that provides higher resolution than traditional karyotyping or FISH (flourescent in-situ hybridization). The Chromosome Analysis Suite software, supplied by the manufacturer, supports analysis of the array results.

The main features of The Cytogenetics 2.7M Array are, unbiased whole genome coverage with the highest density of 2.7 million markers to enable superior resolution. Also, each array includes 400,000 single nucleotide polymorphisms (SNPs) to enable the detection of loss of heterozygosity (LOH), uniparental disomy (UPD), and regions identical-by-descent. All of these are labelled by the software and also links to the UCSC browser, the Toronto variation database (DVG), Online Mendelian Inheritance In Man database (OMIM), and SNP database (dbSNP31).

The arrays are currently used for research purposes only but there is a move towards introducing the technique in the Kuwait Medical Genetic Centre for diagnostic purposes. A previous clinical application of these arrays have reported CNV detection in ID cases (Qiao, Tyson et al. 2013)

In this study, DNA samples of affected individuals have been used for cytogenetic analysis to detect candidate genes associated with ID. As part of the initial plan the selection criteria for samples to be used on the arrays was supposed to be based on the result screening using MLPA in chapter 5. However, due to long delivery delays in receiving the reagents and limited funding, samples for analysis were chosen on the basis of phenotype and family structure. One individual from each family was selected and then an additional affected family member was tested in some families and the results compared across affected within the same family.

6.1.2 Copy number variation (CNV) analysis

With the development of DNA based methods to detect deletions and duplications in the genome, the last two years have witnessed an explosion of interest in human Copy Number Variants (CNVs), defined as DNA deletions or duplications greater than 1 Kb. Deletions and duplications have been recently identified in schizophrenia, mental retardation, autism and epilepsy at loci such as 1q21.1; 2p16.3; 15q11.2 and 15q13.3 (Stefansson, Rujescu et al. 2008; Rujescu and Collier 2009) which affect multiple genes in most cases. Over 100 CNVs have been associated with ID (Webber, Hehir-Kwa et al. 2009) .

However discriminating between benign and pathogenic CNVs is difficult; previously de novo status and size (larger CNVs delete more genes and are therefore more likely to be pathogenic) has been used, but is insufficient to demonstrate pathogenicity. One method to approach this is to analyse families with CNVs and more than one case of ID

(to detect cosegregation with the phenotype), and to select CNVs which can be shown to be under negative genetic selection, using population genetic methods (Stefansson, Rujescu et al. 2008).

These studies rely heavily on new technologies such as cytogenetic arrays to detect CNVs and their status (normal, loss, gain). CNVs can occur in normal individuals as well and hence establishing pathogenicity can be difficult and may require large studies. A key resource for this was provided by Cooper and colleagues, who published a list of potentially pathogenic CNVs based on a study of 15,767 children with various developmental and intellectual disabilities and compared to a CNV map that was produced from 8,329 adult controls (Cooper, Coe et al. 2011).

6.1.3 LOH and blocks of homozygosity

Loss of heterozygosity (LOH) is used by Affymetrix to refer to regions of inherited homozygosity and not limited to a loss through disorders such as cancer. This usage will be carried throughout the thesis. However, it is worth nting that more commonly regions of homozygosity is used (ROH). These regions are common amongst siblings from consanguineous families. It is estimated that 20% of world population live in communities where consanguinity is practiced and preferred with and estimated 8.5% of children being the offspring of such union, most common being that of a first cousn marriage. First cousins share 1/8 of their genes and their off sprint in turn would show autozygosity at 1/16 of all loci, which is referred to as a coefficient of breeding (F) equal to 0.0625 in this case. This likelihood of a rare, recessive disease-causing variants being inherited from a common ancestor via both maternal and paternal lineages, will be more likely compared to the offspring of outbred populations. (Bittles 2008; Musante and Ropers 2014). Tracking these blocks or regions of homozygosity can be of great value when trying to identify recessive mutations as the same alleles or mutation will be inherited. This has yielded great results as reviewed by Al-Kuraya from studies in the Saudi population where consanguinity is very popular and span for generations

(Alkuraya 2010).

6.1.4 Further testing

All techniques require confirmatory and further testing. And a combination of techniques would ensure validated results as well as some regions or mutations that would need to be looked into further either for more resolution or further testing for biological significance that could include animal models. In the case of looking further at regions in question this is usually followed through direct sequencing or in some cases for copy number other methods such as MLPA or real time PCR. With advances in genome sequencing and the drop in cost, it has been the method of choice for follow up and further testing and sometimes even the first method applied, especially with its high resolution and genome coverage that would reduce multiple probe or regions tests to a single run. It has been applied to identify causes of severe ID (Gilissen, Hehir-Kwa et al. 2014) (Musante and Ropers 2014).

Testing for de Novo mutation is possible by genome sequencing affected and parents as done by Rauch and colleague in which 87 de-novo variants were identified in 51 cases from the German Mental Retardation Network (Rauch, Wieczorek et al. 2012). Alkuraya lists 127 novel genes identified, from different studies, by combining Homozygosity mapping with next generation sequencing (Alkuraya 2013) to name one is the identification of a novel mutation of SCARB2 gene in a cases with Progressive myoclonus epilepsy (He, Tang et al. 2014)

6.2 Materials and methods

The study was carried out using Cytogenetics 2.7M arrays, and following the protocol published by Affymetrix in the cyto_assay_usermanual (P/N 702761 Rev. 2). Details of

the materials and methods for the laboratory work can be found in chapter 3. This section only covers methods using the Affymetrix software and follow-up analysis.

6.2.1 Array Data analysis

Data were analysed using the recommended Affymetrix Chromosome Analysis Suite using the analysis parameter file:

Cytogenetics_Arrays.single_sample.def.NA31.v1.chasparam

Reference file for the analysis:

Cytogenetics_Arrays.na31.v1.REF_MODEL

Annotation file:

Cytogenetics_Arrays.na31.annot.db

All these files were accessible through the Affymetrix online documentation.

The software analyses intensity data from the arrays and performs a single sample analysis that compares the results to the reference file and detects the following:

a) Copy number state gain or loss
b) Mosaicism, which are regions of non-integer copy number gain or loss as a CN? status between 1 and 3 (these were not looked into in this study).
c) Loss of heterozygosity (LOH), which are regions where most SNPs do not display heterozygosity
d) Long contiguous stretches of homozygosity (LCSH), which is the loss of heterozygosity in regions of normal expected copy number.

It is possible to create a reference file that is population specific but this was not done in this study due to limited resources. More array chips would be required to construct a reference file for the Kuwaiti population.

6.2.1.1 Quality control for Arrays

Quality control (QC) is a very important aspect of any experimental design that measures precision that affects the reliability of results. This is even more sensitive in a diagnostic laboratory, where patients await their results and reporting false positive or negative results as genuine must be avoided.

The QC measures can cover various areas of the experiment, from design, to laboratory setting to software QC parameters. For example including negative or positive control samples can give an indication about the reliability of a method to detect known variants. Qiao and colleagues used cases with known CNVs in order to check the reliability of the arrays in detecting them (Qiao, Tyson et al. 2013). Other measures include running duplicate samples to check for human handling errors or even imprecision is handling stations such as the washing stations for the array chips. In terms of experimental design, running arrays on the same day would reduce variations that could be due to sample handling and other conditions such as temperature and solution mixing and incubation time.

Quality control checks and parameters in the software were used to filter samples before the data were normalised, as recommended by Affymetrix, these include:

1- SNPQC (SNP Quality Control: how well the A and B allele can be resolved), the recommended value being ≥1.1
2- MAPD (Median Absolute Pair- wise Difference: how similar the signal distribution of the sample is relative to the reference model file), the recommended value being ≤0.27
3- Waviness-seg-count (number of autosomal segments using a 50 marker kb resolution), the recommended value being ≤30
4- Median raw intensity (median pre-normalization signal intensity of genomic

probes: a measure of brightness), the recommended value being ≥2000

5- Antigenomic ratio (median of antigenomic probe intensities divided by the median of copy number probe intensities: a measure of background), the recommended value being ≤0.27.

6.2.1.2 Results display and selection criteria

Results were viewed with focus on two variation types, CNVs and LOH, each list was selected for viewing separately on the Chromosomal suite software and the data was exported in a text file, which was then imported into Excel. A snap shot of the genomic view was taken using the clip capture capability of Microsoft office OneNote. This was also used to snap an image of the detailed view window for the shared regions of gains or loss **Figure 50**.

For illustrative purposes, below in **Figure 50** is a snap shot of the genomic view of the array showing regions of gains and losses per chromosome as analysed by the Chromosome Analysis Suite. Red arrows indicate regions of loss, while regions of gains are indicated in blue.

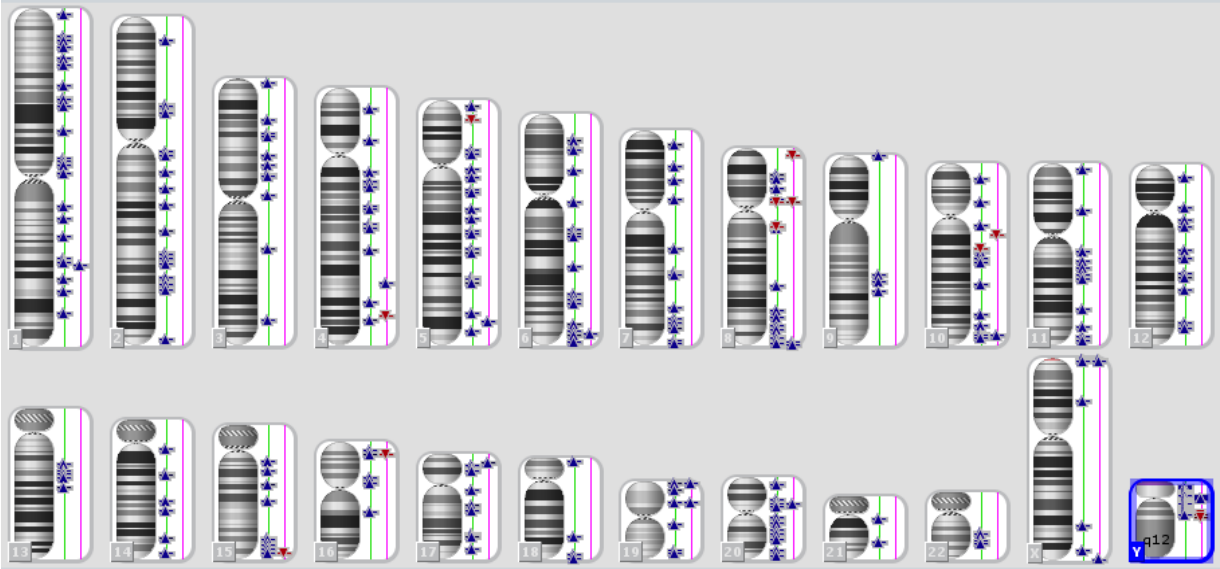


Figure 50 Genomic view of Family 2 (array 2.1 on the right (green line) and array 2.2 on the left (red line)), with loss indicated in red and gain in blue. Some gains or losses could be artefacts and cause by DNA quality and technical problems.

Variation was accepted for filtration at 85% confidence level.

6.2.1.3 CNV analysis

Following recommendations by Qiao and Miller (Qiao, Tyson et al. 2013) CNVs were filtered by the following and at each filtration level results were counted via the pivot table function in excel:

For each individual cases filtration was as follows

a) Loss and gain CNVs were separated into different list based on size with size cut off being 200kbp for losses and 500kbp for gains.
b) For each list results were filtered further to select: i) CNV regions which cover genes and 50 or more markers, ii) CNV regions which cover genes, 50 or more markers, and no more than 2 reported studies in the Database of Genomic Variations.

c) The final filtered results were looked at further using the software to check if regions are genuine by observing the marker coverage and log2ratio graphs. Illustrated in the results section.
d) Each filtered list was labeled and stored for further future research and to build a database of results observed.

As a note if more resources were available additional runs would have been made and further testing would have been carried out to identify potentially pathogenic regions. CNVs were looked at in individual, then as shared by siblings and finally across all cases.

6.2.1.4 LOH analysis

Similar to CNVs filters were applied to narrow down the LOH regions. For each individual the LOH percentage of the human genome was calculated using the kbp size of the LOH regions combined for the individual out of the human genome 19 build estimated as 3,156,105 kbp in size. In addition, the number of LOH regions covering genes, OMIM entries, and the key words mental retardation were reported in the results for each individual. Furthermore, the LOH regions covering disorders listed on OMIM with the key words "mental retardation" were compared across all families and summarised in the results.

6.2.3 Further testing using exome sequencing

Two affected siblings from family 7 were tested further in order to identify potentially pathogenic variation that contributed to the disorder using exome sequencing. The exome method employed was as follows:

Target enrichment of the exome has been performed using the SureSelect Human All Exon Kit (50Mb). Sequencing has been done using the Illumina Genome Analyzer IIx.

Results were received in an excel format. The files were generated from sequence data aligned to the hg18 reference genome with Novoalign, variants called with SAM tools and annotated via multiple passes through annovar. The results included, Sample ID,

Chromosome, chromosomal start and end positions, reference allele, alternative allele, zygosity of the variant call, exonic or splicing (10bp), gene in which the variant resides, cosequence of coding variant (synonymous, nonsynonymous, stop gain, stop loss, frameshift deletion, frameshift insertion , nonframeshift deletion, nonframeshift insertion), and Novelty (not in dbSNP129,dbSNP132, 1000genomes or our inhouse database), precompiled SIFT score if available, dbSNP129 ID, and dbSNP132 ID.

A log2ratio plot was also received for each individual compared to a healthy control reference sample and it is included in the results section below.

6.2.3.1 Narrowing down exome results

Exome results were used as further testing to identify potentially pathogenic single nucleotide variation (SNV). The assumption was that a homozygotic variation under a recessive form of inheritance maybe be residing within LOH regions and shared amongst the affected siblings. This was the basis of the selection criteria in which shared LOH regions (shared amongst sibs and not unaffected parent) from array results were compared to exome results, SNVs that were homozygous and that were not synonymous were selected and the last selection criteria took into account publish Cooper list of potentially pathogenic genes associated with ID.

6.3 Results

Results are divided into five parts: DNA concentration and array labeling, overall variations summary, CNV analysis, LOH analysis and the results of further testing using exome sequencing.

6.3.1 DNA concentrations and array labels

The initial concentration and after the amplification and purification steps are listed below for the 18 samples that generated results successfully out of the 24 samples run, these are listed below in **Table 41**

Table 41 Sample information and QC metrics (family number, sample label, initial sample concentration and the concentration after amplification and purification, gender)

		Sample concentration & OD reading					Concentration & OD reading after amplification & purification					
family	Array #	ng/ul	A260	A280	260/280	260/230	ng/ul	A260	A280	260/280	260/230	Gender
1	100824-01	287.91	5.758	3.176	1.81	1.97	1491.35	29.827	15.812	1.89	2.35	m
1	10120101	292.43	5.849	3.236	1.81	1.94	1398.65	27.973	14.827	1.89	2.32	m
2	100824-02	52.23	1.045	0.535	1.95	1.22	1176.53	23.531	12.439	1.89	2.35	m
2	10120602	77.29	1.546	0.784	1.97	1.44	1208.58	24.172	13.214	1.83	2.26	m
3	B0921	51.48	1.03	0.529	1.95	2.03	1102.94	22.06	12.26	1.8	2.23	m
3	10120102	37.32	0.746	0.383	1.95	1.22	1349.72	26.994	14.318	1.89	2.33	m
4	C0921	641.92	12.838	6.849	1.87	2.16	1184.24	23.69	12.67	1.87	2.32	m
5	D0921	91.84	1.837	0.955	1.92	1.5	1165.46	23.31	12.95	1.8	2.25	f
6	10092802	285.52	5.71	3.033	1.88	2.06	1121.82	22.436	12.168	1.84	2.35	f
7	10100601	368.46	7.369	4.01	1.84	2.09	1406.25	28.125	15.006	1.87	2.32	f
7	10100602	115.77	2.315	1.261	1.84	1.75	1414.16	28.283	15.006	1.88	2.35	f
7	10101102	283.84	7.677	4.19	1.83	2.1	1277.01	25.54	13.804	1.85	2.32	f
8	10101101	381.16	7.623	3.951	1.93	1.08	1252.65	25.053	13.469	1.86	2.33	m
9	10121501	115.82	2.316	1.471	1.58	1.2	1249.23	24.985	13.395	1.87	2.35	m
9	10121502	30.83	0.617	0.353	1.75	1.4	1385.18	27.704	14.844	1.87	2.31	m
10	10120801	125.32	2.506	1.382	1.382	1.81	1297	25.94	14.638	1.77	2.24	m
rsts 1	10191003	1322.2	26.44	14.2	1.86	1.96	1211.58	24.232	12.837	1.89	2.35	f
rsts 1	10120601	265.23	5.305	2.798	1.9	1.77	1307.67	26.153	14.113	1.85	2.29	m

6.3.2 Variations summary

The variations found were classified by type for each family and summarized in **Table 42** next page and graphically shown in **Figure 51**. The majority of results were that of LCSH (long continuous stretches of homozygosity) regions followed by regions of gain, then LOH regions, then regions of loss and finally few mosaic regions in half of the arrays.

The average is also listed for each type but not taking into the calculation samples 9.1, 9.2, 10.1 and A.1 as they would distort the average and had too many variations than expected.

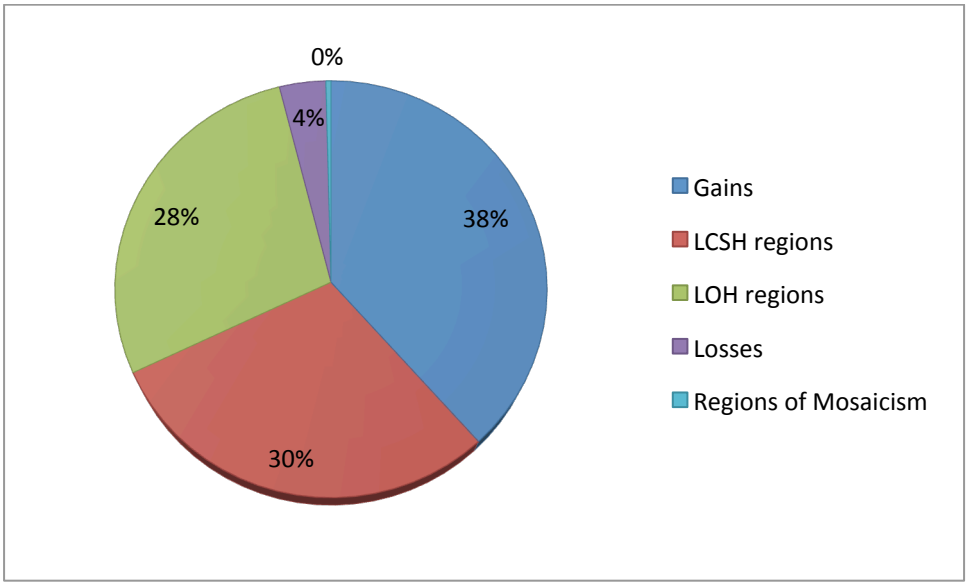


Figure 51: showing the proportion of each variation type found in all of the array results

Table 42 showing the summary of variations found of each type for each array with the overall total and average

Sample number	1.1	1.2	2.1	2.2	3.1	3.2	4.1	5.1	6.1	7.1	7.2	7.3	8.1	9.1	9.2	10.1	A.1	A.2	Total	Average
gender	M	M	M	F	M	M	M	F	F	F	F	F	F	M	M	M	F	M	/	/
affected status (y/n)	y	y	y	y	y	y	y	y	y	y	n	y	y	n	y	y	y	y	/	/
Total variations	111	243	303	365	368	220	219	139	487	150	347	123	264	4910	1634	737	174	728	11522	195.2
Total variations at 85% confidence	95	154	216	351	298	153	176	103	419	98	265	100	174	4552	1492	621	166	655	10088	153.8
Gains	30	20	16	281	96	5	34	20	280	41	106	33	27	1058	190	491	74	404	3206	59.1
LCSH regions	28	62	95	38	97	70	70	40	70	28	76	30	71	166	187	59	43	148	1378	45.4
LOH regions	28	62	98	27	91	70	65	39	55	28	73	30	71	104	99	55	40	93	1128	43.2
Losses	9	10	7	4	11	8	7	4	11	0	8	6	5	3211	1016	11	9	10	4347	5.5
Regions of Mosaicism	0	0	0	1	3	0	0	0	3	1	2	1	0	13	0	5	0	0	29	0.6

6.3.3 CNV analysis results

CNV analysis results were filtered as per the method mentioned in the materials and methods section in this chapter. The results are summarized in two tables below one for losses and the other for gains. No losses were observed for individual 7.1. After applying different selection criteria the remaining number of losses or gains dropped significantly.

As evident in the two tables below samples from individuals 9.1 , 9.2 and to some extent A.2 had far too many CNVs than the average reported. The final selected results for loss or gain

and in the different size range (selection criteria being covering 50 markers or more, covering genes, and with no more than 2 reported studies in DGV), were looked at closer on the detailed view option on the software. This included looking at marker coverage and log2ratio plot as illustrated on **Figure 52** below.

Table 43: the number of losses found in each individual sample with various selection filters applied

Sample number	1.1	1.2	2.1	2.2	3.1	3.2	4.1	5.1	6.1	7.1	7.2	7.3	8.1	9.1	9.2	10.1	A.1	A.2	Total	Average
gender	M	M	M	F	M	M	M	F	F	F	F	F	F	M	M	M	F	M	/	/
affected status (y/n)	y	y	y	y	y	y	y	y	y	y	n	y	y	n	y	y	y	y	/	/
Losses	9	10	7	4	11	8	7	4	11	0	8	6	5	3211	1015	11	9	10	4346	7.3
losses > 200kbp	0	0	0	1	1	0	2	1	0	0	1	0	0	29	11	1	0	0	47	0.4
losses > 200kbp covering gene and 50 or more markers	*0*	*0*	*0*	*1*	*0*	*0*	*2*	*0*	*0*	*0*	*0*	*0*	*0*	*23*	*6*	*0*	*0*	*0*	32	0.2
losses > 200kbp covering gene, covering 50 or more markers, and not >2 reported studies in DGV	0	0	0	1	0	0	0	0	0	0	0	0	0	8	2	0	0	0	11	0.1
losses < 200kbp	9	10	7	3	10	8	5	3	11	0	7	6	5	3182	1005	10	9	10	4300	6.9
losses < 200kbp, covering 50 or more markers, covering genes	1	2	1	0	0	0	0	0	0	0	1	0	0	488	288	1	0	1	783	0.4
losses < 200kbp, covering 50 or more markers, covering genes, and not > 2 reported studies in DGV	0	0	0	0	0	0	0	0	0	0	0	0	0	133	56	0	0	1	190	0.1

Table 44 the number of gains found in each individual sample with various selection filters applied

Sample number	1.1	1.2	2.1	2.2	3.1	3.2	4.1	5.1	6.1	7.1	7.2	7.3	8.1	9.1	9.2	10.1	A.1	A.2	Total	Average
gender	M	M	M	F	M	M	M	F	F	F	F	F	F	M	M	M	F	M	/	/
affected status (y/n)	y	y	y	y	y	y	y	y	y	y	n	y	y	n	y	y	y	y	/	/
gains	30	20	16	283	96	6	37	20	280	42	110	36	27	1058	191	493	75	405		
gains >500kbp	4	5	2	8	15	1	2	2	16	7	11	6	2	2	2	24	7	10	126	6.5
gains > 500kbp covering genes and 50 or more markers	*4*	*3*	*2*	*6*	*10*	*1*	*0*	*1*	*12*	*5*	*6*	*4*	*1*	*0*	*0*	*21*	*5*	*6*	87	4.4
gains > 500kbp covering genes, covering 50 or more markers, and not >2 reported studies in DGV	0	0	0	1	0	0	0	1	1	1	1	2	0	0	0	0	1	0	8	0.5
gains< 500kbp	26	15	14	273	81	4	32	18	264	34	95	27	25	1056	188	467	67	394	3080	91.3
gains < 500kbp, covering 50 or more markers, covering genes	8	2	3	127	20	0	5	1	56	6	28	2	1	385	110	101	14	138	1007	27.4
gains < 500kbp, covering 50 or more markers, covering genes and not > 2 reported studies in DGV	1	0	0	0	0	0	0	0	0	0	0	0	0	362	233	0	0	1	597	0.1

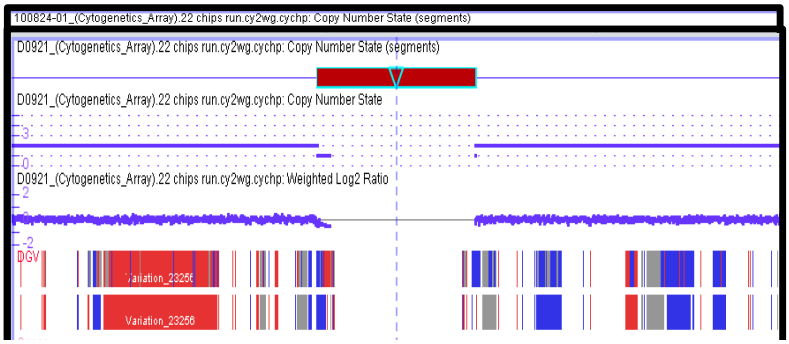


Figure 52: showing the detailed view of two CNV regions and different plots and database alignments. Top screen shows two regions of gain followed by Copy number status, Log2 Ratio, DGV entries, gene list, OMIM entries. The bottom one shows a deletion that aligns to the centromere with no marker coverage.

No CNVs showed significant association with ID and without further testing it would be difficult to go back on the selection criteria and look at more regions without being able to carry out further investigation. The data was kept and labelled according to the criteria for future testing. CNVs were looked at within family and across all families but candidate regions were found that relate to ID.

6.3.4 LOH analysis results

For LOH regions there were many found as illustrated in **Figure 53** below. **Table 45** summarises the results at each filtration criteria applied. Regions that overlapped reported disorders with ID (mental retardation was used as the key ward as it is still used in most disorder names) on OMIM were compared across families. These were listed in **Tables 46** and **47**, were 47 regions were identified shared in at least two individuals and a total of only 14 regions were share amongst sibs some in family 3 and others in family A. Very high level of LOH percentages were found which might suggest further relations and more common genes. The regions were filed for further testing when resources are available.

However, further testing was available for one pair of affected sibs in family 7 and one of the unaffected parents was tested on the array as well. In this case the LOH results were looked into further and the regions that were shared in sibs and not parent were listed in the next section.

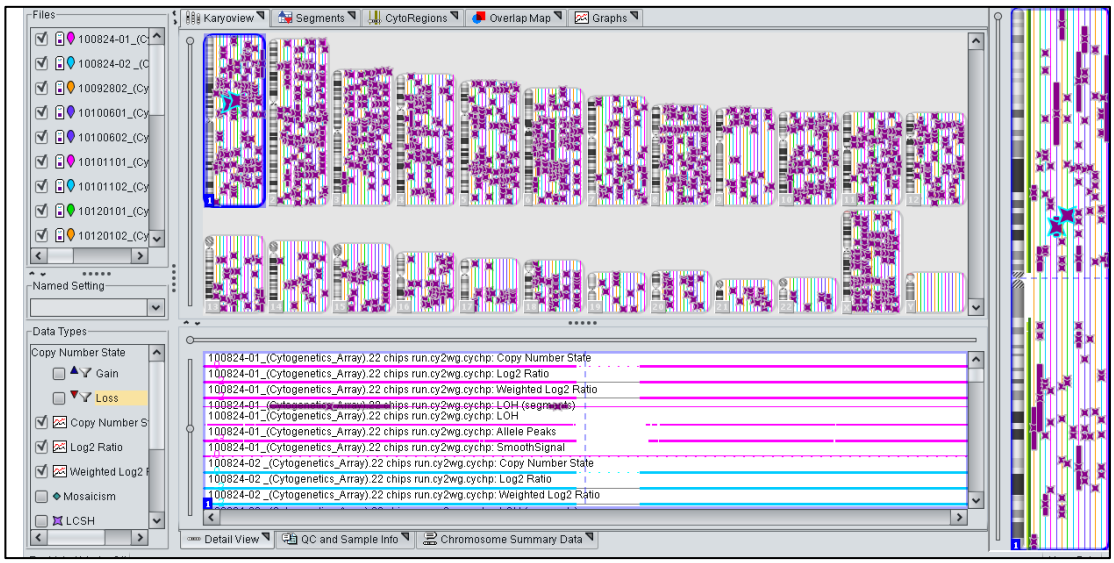


Figure 53: showing the LOH regions observed in 15 individuals aligned across all chromosomes

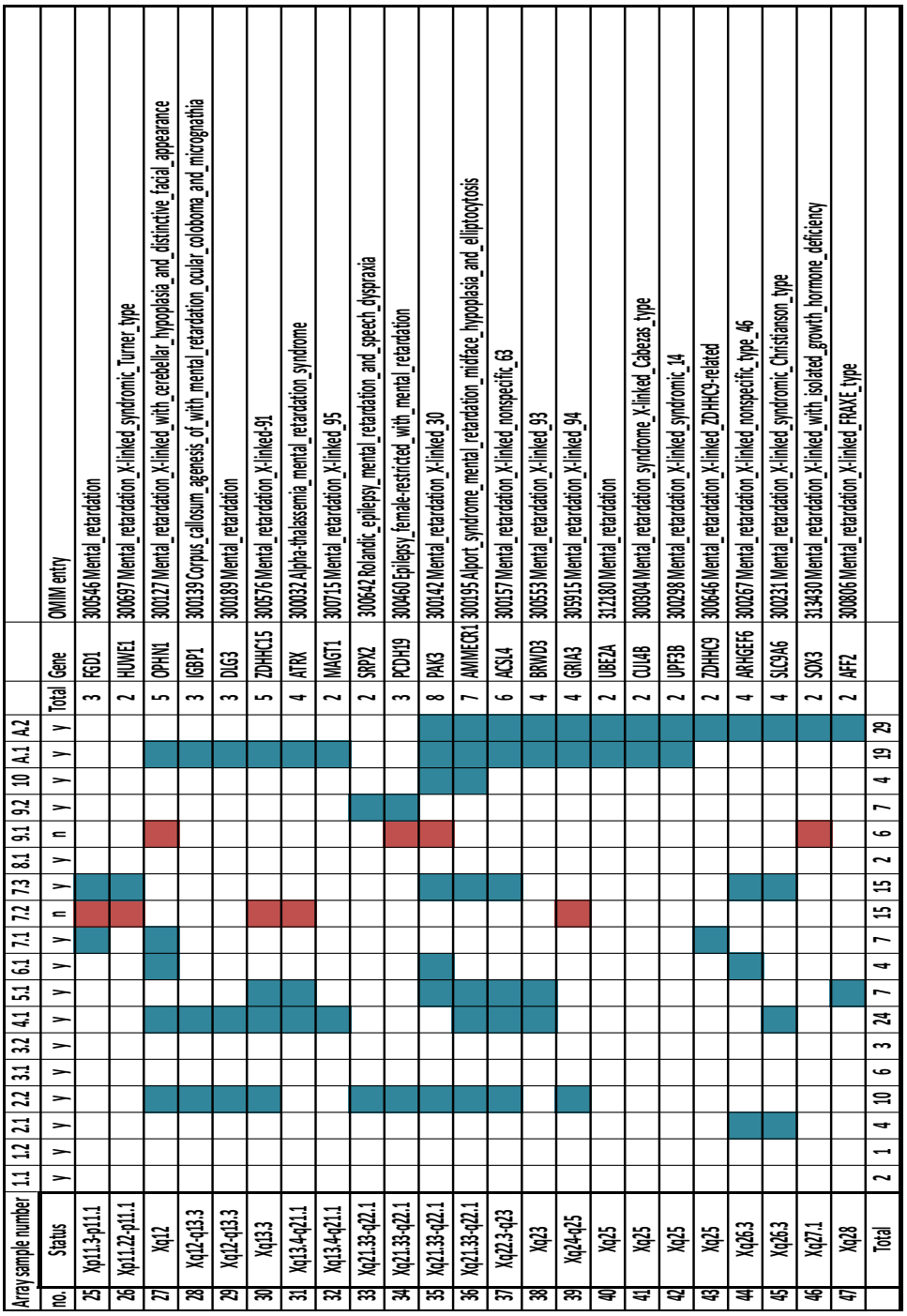
Table 45 the number of LOH regions observed in each individual at different selection filtered applied

Sample number	1.1	1.2	2.1	2.2	3.1	3.2	4.1	5.1	6.1	7.1	7.2	7.3	8.1	9.1	9.2	10.1	A.1	A.2	Total
gender	M	M	M	F	M	M	M	F	F	F	F	F	F	M	M	M	F	M	/
affected status (y/n)	y	y	y	y	y	y	y	y	y	y	n	y	y	n	y	y	y	y	/
number of LOH regions	35	107	141	27	121	103	84	57	82	53	110	40	116	200	146	101	42	94	1659
number of LOH regions at 85% confidence	28	60	96	27	88	69	63	39	54	26	73	29	67	101	96	51	40	93	1100
% of LOH in the genome	9.56	5.75	5.98	8.58	15.2	8.54	21.3	10.8	6.19	6.72	7.25	7.34	6.72	4.34	18.2	2.53	22.4	36.6	/
number of LOH regions covering genes	28	54	90	27	83	64	62	38	51	25	68	29	59	87	86	47	40	93	1031
LOH regions covering genes that are listed on OMIM	27	27	54	26	65	45	53	29	36	23	43	25	40	41	49	31	32	89	735
LOH regions covering genes that are listed on OMIM with key words "Mental retardation" in the tiltle	2	1	3	3	8	2	11	4	4	5	9	5	3	7	5	3	9	10	94

Table 46 showing the presence/absence of 24 LOH regions in each individual, chromosomal region, gene covered and it's relevant OMIM entry that has the key words "mental retardation". The affected individuals were marked in green while unaffected ones marked in red.

Array sample number		1.1	1.2	2.1	2.2	3.1	3.2	4.1	5.1	6.1	7.1	7.2	7.3	8.1	9.1	9.2	10	A.1	A.2			
no.	Status	y	y	y	y	y	y	y	y	y	y	n	y	y	n	y	y	y	y	Total	Gene	OMIM entry
1	2q22.3-q23.1																			6	MBD5	611472 Mental_retardation_autosomal_dominant_1
2	2q33.1																			7	SATB2	608148 Cleft_palate_and_mental_retardation
3	3p26.3-p26.1																			3	CRBN	610262 Mental_retardation_autosomal_recessive_2A
4	4q26-q28.1																			2	PRSS12	606709 Mental_retardation_autosomal_recessive_1
5	5p15.2																			6	CTNND2	604275 Mental_retardation_in_cri-du-chat_syndrome
6	6q16.1-q21																			2	GRIK2	138244 Mental_retardation_autosomal_recessive_6
7	8q12.1-q12.2																			4	CA8	114815 Cerebellar_ataxia_and_mental_retardation_with_or_without_quadrupedal_locomotion_3
8	11q24.2-q24.3																			3	KIRREL3	607761 Mental_retardation_autosomal_dominant_4
9	12q13.11-14.1																			2	DIP2B	611379 Mental_retardation_FRA12A_type
10	16q11.2-q23.1																			2	CDH3	114019 Mental_retardation_autosomal_dominant_3
11	Xp22.13-p22.12																			5	RPS6KA3	300075 Mental_retardation_X-linked_nonspecific
12	Xp22.12-p22.11																			3	SMS	300105 Mental_retardation_X-linked_Snyder-Robinson_type
13	Xp22.32-p21.3																			2	ARX	300382 Epilepsy_myoclonic_with_mental_retardation_and_spasticity/300382 Mental_retardation_X-linked_36_43_54
14	Xp21.3-p21.2																			3	ILIRAPL1	300206 Mental_retardation_X-linked_21_34
15	Xp11.3-p11.22																			3	ZNF41	314995 Mental_retardation
16	Xp11.3-p11.22																			2	SYP	313475 Mental_retardation_X-linked_with_or_without_epilepsy
17	Xp11.4																			2	ATP6AP2	300556 Mental_retardation_X-linked_with_epilepsy
18	Xp11.23-p11.22																			6	SHROOM4	300579 Stocco_dos_Santos_X-linked_mental_retardation_syndrome
19	Xp11.22																			3	FTSJ1	300499 Mental_retardation_X-linked-9
20	Xp11.22																			4	HSD17B10	300256 Mental_retardation_X-linked_17_31_microduplication
21	Xp11.22																			3	KDM5C	314690 Mental_retardation_X-linked_syndromic_JARID1C-related
22	Xp11.22																			3	ZNF81	314998 Mental_retardation_X-linked_45
23	Xp11.22																			3	ZNF674	300573 Mental_retardation
24	Xp11.3-p11.1																			4	PHF8	300560 Mental_retardation_syndrome_X-linked_Siderius_type

Table 47 showing the presence/absence of further 23 LOH regions in each individual, chromosomal region, gene covered and it's relevant OMIM entry that has the key words "mental retardation". The affected individuals were marked in green while unaffected ones marked in red.



Array sample number		1.1	1.2	2.1	2.2	3.1	3.2	4.1	5.1	6.1	7.1	7.2	7.3	8.1	9.1	9.2	10	A.1	A.2			
no.	Status	y	y	y	y	y	y	y	y	y	y	n	y	y	n	y	y	y	y	Total	Gene	OMIM entry
25	Xp11.3-p11.1																			3	FGD1	300546 Mental_retardation
26	Xp11.22-p11.1																			2	HUWE1	300697 Mental_retardation_X-linked_syndromic_Turner_type
27	Xq12																			5	OPHN1	300127 Mental_retardation_X-linked_with_cerebellar_hypoplasia_and_distinctive_facial_appearance
28	Xq12-q13.3																			3	IGBP1	300139 Corpus_callosum_agenesis_of_with_mental_retardation_ocular_coloboma_and_micrognathia
29	Xq12-q13.3																			3	DLG3	300189 Mental_retardation
30	Xq13.3																			5	ZDHHC15	300576 Mental_retardation_X-linked-91
31	Xq13.4-q21.1																			4	ATRX	300032 Alpha-thalassemia_mental_retardation_syndrome
32	Xq13.4-q21.1																			2	MAGT1	300715 Mental_retardation_X-linked_95
33	Xq21.33-q22.1																			2	SRPX2	300642 Rolandic_epilepsy_mental_retardation_and_speech_dyspraxia
34	Xq21.33-q22.1																			3	PCDH19	300460 Epilepsy_female-restricted_with_mental_retardation
35	Xq21.33-q22.1																			8	PAK3	300142 Mental_retardation_X-linked_30
36	Xq21.33-q22.1																			7	AMMECR1	300195 Alport_syndrome_mental_retardation_midface_hypoplasia_and_elliptocytosis
37	Xq22.3-q23																			6	ACSL4	300157 Mental_retardation_X-linked_nonspecific_63
38	Xq23																			4	BRWD3	300553 Mental_retardation_X-linked_93
39	Xq24-q25																			4	GRIA3	305915 Mental_retardation_X-linked_94
40	Xq25																			2	UBE2A	312180 Mental_retardation
41	Xq25																			2	CUL4B	300304 Mental_retardation_syndrome_X-linked_Cabezas_type
42	Xq25																			2	UPF3B	300298 Mental_retardation_X-linked_syndromic_14
43	Xq25																			2	ZDHHC9	300646 Mental_retardation_X-linked_ZDHHC9-related
44	Xq26.3																			4	ARHGEF6	300267 Mental_retardation_X-linked_nonspecific_type_46
45	Xq26.3																			4	SLC9A6	300231 Mental_retardation_X-linked_syndromic_Christianson_type
46	Xq27.1																			2	SOX3	313430 Mental_retardation_X-linked_with_isolated_growth_hormone_deficiency
47	Xq28																			2	AFF2	300806 Mental_retardation_X-linked_FRAXE_type
	Total	2	1	4	10	6	3	24	7	4	7	15	15	2	6	7	4	19	29			

6.3.5 Further testing results

Exome sequencing results were received in two formats for individual 7.1 and 7.3. One was of a log 2 ration graphs for all chromosomes, shown in **Figure 54**, and the second was a table listing the specific single nucleotide variations found, a summary of the zygosity status and consequence of each SNV is listed in **Table 48** below.

Regions of homozygosity were looked for further analysis and to compare it with the SNV results. This yielded in 13 regions of varying size out of an initial 25 regions that encompass many genes and were located on chromosomes 1, 8, 9, 11, 12, 14, 16, 19, 21 and X. These were shared amongst the sibs but not the mother (data not shown). Several of these regions cover genes associated with syndromes that have ID reported in at least one individual with this condition. These include: Holt-Oram syndrome, Uinar mammary syndrome, Wrinkly skin syndrome, Townes-Brocks branchiootorenal like syndrome, Opitz G syndrome, X-linked mental retardation and Fragile X (OMIM results in the appendix).

The DNA samples of the affected sibs were sent for exome sequencing and data analysis. When comparing the exome sequencing results, and after applying several filtering criteria Three variations were found including two point mutations both within one of the homozygosity blocks that have also covered the genes listed in the Cooper list.

These corresponded to single nucleotide variants at a homozygous status. One thought to be non-synonymous in gene QPRT, The chromosomal coordinate being Chromosome 16:29615851, with a nucleotide variation from the expected adenine (A) into guanine (G). The second variation affecting splicing in gene GDPD3, the chromosomal coordinate being Chromosome 16:30027026 with variation of cytosine (C) to adenine (A). Searching the SNP references, supplied with the exome analysis (rs9932770, rs8061772), in the database of SNP (dbSNP), the clinical significance was reported as not applicable and clinical channel unknown, which comes from the 1000 genome project. Searching the results on PolyPhen-2 tool (for prediction of functional effects of human nsSNPs), the first SNP was predicted to be benign, while the second one failed to generate results within the tool, Perhaps because it is a splice site variant, and therefore not in the coding DNA.

Figure 54: Log 2 ration plots of SNV results of individual 7.1 on the left and 7.3 on the right each compared to a normal control.

Table 48 the list of SNVs found in each individual including zygosity state and consequence of each variation.

SNV type	7.1	7.3	Grand Total
HET	9646	9836	19482
frameshift deletion	17	19	36
frameshift insertion	14	10	24
nonframeshift deletion	42	50	92
nonframeshift insertion	25	26	51
nonsynonymous SNV	4919	4940	9859
splicing	926	944	1870
stopgain SNV	54	57	111
stoploss SNV	11	10	21
synonymous SNV	3638	3780	7418
HOM	6738	6548	13286
frameshift deletion	24	29	53
frameshift insertion	49	48	97
nonframeshift deletion	22	20	42
nonframeshift insertion	27	26	53
nonsynonymous SNV	3419	3308	6727
splicing	623	626	1249
stopgain SNV	11	9	20
stoploss SNV	13	15	28
synonymous SNV	2550	2467	5017
Grand Total	16384	16384	32768

Table 49 showing the SNVs found and the selection criteria applied to arrived at the potentially pathogenic SNVs.

Filter	7.1	7.3	Total
Total SNV results from exome sequencing	**16384**	**16384**	**32768**
Total SNVs overlapped with 16 LOH region	**552**	**563**	**1115**
Homologous SNVs	**396**	**398**	**794**
SNVs Shared amongst sibs	**368**		
Without synonymous mutations	**221**		
SNVs overlap with Cooper reported genes	**2**		

Furthermore from searches on the dbSNP the database it was found that rs9932770 has a frequency of approximately 0.001. Carriers have been observed in the HapMap East Asian, Indian and African populations, but not in populations of European ancestry. This makes it very interesting. In contrast, rs8061772 is much more common, with a combined frequency of 0.111 across populations hence, it is seen in most populations world wide so it is unlikely to be a functional mutation. It can be seen from **Table 49** how the different selection criteria affecting the total SNVs from 16384 to 2 only.

6.3.5 Other results

Specific searched were made in the array results. One was of SHOX gene as it was in the MLPA kit as was searched for comparison. It was found as part of CNVs at a gain status in samples 1.1, 3.1, 6.1, 7.1, and 7.3.

The other for Rubinstein Taybi in which a gain was found in individual A.2 at EP300 gene, which is associated with RTS and mutations in it account for 3% of cases. Without the specific searches these would not maked it to the final filtered result due to the last criteria of not being included in a result reported in the DVG.

The results were compared with Cooper list of potentially pathogenic CNVs without the final filter, which restricted DVG reported studies to 2 and reduced the number of CNVs significantly. These are covered further in the discussion.

6.4 Discussion

In terms of the success rate of the samples run, out of the 24 arrays that were run, only 18 produced results that were analysed with the software. Arrays could have failed due to a number of reasons including DNA quality, error in carrying out procedures, or the fact that some reagents were used beyond their expiry date. Also, one wash station

had a malfunction and was later repaired resulting in the failure of at least 3 arrays that were attached to that station. DNA clean-up procedure can be applied to improve on the quality of the DNA as recommended in the Affymetrix user manual.

6.4.1 Family results summary

Looking at each family, for family 1, two gains were found to be shared amongst both affected sibs, one covering 2 genes and the other 42 genes, and both being on chromosome 8. Both of the regions were also shared with other families and hence are more interesting to follow up but it could be non pathogenic and hence testing the parents would give more insight on potential pathogenicity.

For family 2, a gain on chromosome 1 covered 14 genes that have been reported in the Cooper list although the CN status differed. In terms of shared regions, two gain CNVs were found which also had an overlap with the shared regions identified for family 1. The shared regions would be of more interest to investigate first.

For family 3, two gains were found shared amongst the affected sibs as well as those of families 1 and 2 on chromosome 8. In additions, 3 gain CNVs covering 126 genes reported as a loss in the Cooper list. In addition SHOX gene can be investigated further in this family as potentially causative gene as it although it appeared t be at a loss using MLPA technique for individual 3.1.

For family 4, only one affected individual was sampled, with all falters applied, no results would be reported, reducing filters (marker count and DGV reported studies) would yield 29 regions of loss and gain, 3 of which being regions of loss and the remaining gain regions. Of interest would be the following genes: CSRNP3, SLC2oA2, TSNARE1, ADAM12 (which also appeared in family 3), EID3, TSHZ2, as they do not have previously reported variations available on the DGV database (data not shown).

For family 5, which again had only one affected member looked at, a total of 13 regions of gain and one of loss were found, spanning 5 chromosomes (17, 19, 20, 22 & Y). From

the autosomes there are 3 genes of interest PITPNM3, CDH4 and MTID3 as they do not appear to have previously reported variations database (data not shown).

For family 6, although a single sample was used a loss was identified on chromosome 2 that covers NRXN1 gene associated to Pitt-Hopkins-like syndrome 2 and Autism susceptibility (based on OMIM results). This would be missed in the very stringent criteria. In addition, a gain was found that is shared with family 7 and 9, on chromosome 1 that covers 4 genes reported in the Cooper list as a loss.

For family 7, several homozygosity regions are of interest, that overlap with OMIM listed syndrome related to ID namely X-linked mental retardation and fragile X syndrome that would require follow up as well as the two SNV reported by exome sequencing.

For family 8, again a single sample was run and no CNVs when applying all filters, while reducing them revealed 20 regions of gain and 2 regions of loss. The regions of loss overlap C9orf3 gene on chromosome 9 (with no previously reported variation) and PRKG1 on chromosome 10. Other genes of interest in gain regions would be SDC3, TCF7, MAN1A1, C9orf3, RASGEF1A, PITPNM3, and CDH4 (the last two also were found in family 5, data not shown).

For family 9, a total of 8 combined losses were found all on chromosome X (these losses were narrowed down by having corresponding OMIM entries), One gain was found on the Y chromosome and it was also shared amongst both sibs, 3 losses were found on chromosomes 2 and 3 and shared amongst the sibs with no corresponding OMIM entries, and shared gains on chromosome 10. Also, three CNVs were found to overlap genes listed in the Cooper list, one of which matched the CN status reported and was share amongst the sibs, localized on chromosome 2 and covering gene SEPT2. This would be interesting to investigate first.

For family 10, 3 gains were found that overlapped genes found in the Cooper list in a different CN state.

For family A, which is of RTS individuals, the main finding in this family was that of a gain in a region covering EP300 gene which is one of two genes known to be associated with RTS disorder. It was only found in one individual but it is work investigating again as it might have been missed in the array analysis in the other sib and would require further confirmatory testing. Also a homozygous loss was found in one individual on chromosome 7 that covers CNTNAP2 gene associate to Autism susceptibility and Pitt-Hopkins like syndrome 1. Three regions of gain were found to be shared, two on chromosome 8, with families 1, 2, and 3. On one individual a gain on chromosome was found encompassing CHAF1A gene found on the Cooper list with the same CN status.

Overall LOH regions can be investigated further especially those overlaying ID related disorders. Some showed more LOH percentage than others, which might account to intermarriages that span few generations back which is very common in this region.

6.4.2 Limitations

There were several limitations in this study ranging from experimental design, supplies, reference sample used, technical support and database used.

The quality control measures were poor in this study due to restriction in the budget and the arrays supplied. No positive controls were tested as compared to the study by Qiao and colleagues, which is partly also due to not having a sample with detected CNV associated with ID from the Kuwaiti population, although any known CNV or even larger aberration would have been useful in establishing aberration detection success. Also running duplicates can give indications on quality control.

It is also worth mentioning that the reference file used was based on the software data and not on a run sample that would take into account laboratory conditions, and may also reduce variation due to ethnic differences.

The arrays produced many results, which made interpretation overwhelming. Different filtering criteria were applied to identify potentially pathogenic CNVs. Some arrays generated more variation results than others. However, establishing pathogenicity of these is difficult with the current experimental design.

As further testing of family members would be vital to establish possible pathogenicity and association with a disorder. Since more than 300 genes are associated with ID it is more difficult to draw conclusions at the scale of this study.

6.4.3 Conclusion and future work

Before going further and testing other family members to establish link to disorder it is important to have a confirmatory test of the CNV which is usually done by several other methods including MLPA, real time PCR and array CGH.

There are several interesting regions to follow up on but none are conclusive without further testing. Having more individuals from the same family was very useful for narrowing down the results. However, this does not mean that the remaining results are to be dismissed completely as there maybe CNVs that had no corresponding entries or were not shared that might still be pathogenic. Even in the case of exome sequencing other mutations could be pathogenic but these were not looked into in the scope of this study but could be explored further in the future.

For follow up analysis, we would prioritise CNVs with deletions that are associated with ID related disorders, then those that share the CN status with those in the Cooper list, followed by CNVs that are shared within families and the two shared across 4 families. In addition to the two SNVs for family 7 were identified with one potentially being the causative mutation.

It is important to note that there are other possible causes of pathogenicity are possible that might not be detected in this method that includes point mutations, and epigenetic factors such as methylation. Also SNP data was generated by the arrays,

which has not been analysed here because the sample size was too small to have power to identify ID-associated SNPs.

The underlying hypothesis was that an autosomal recessive variation was the cause of ID in each family and this has driven the focus of the investigation. Cases where only one individual had the variation do not fit with this hypothesis but could be genuine in the case of RTS where the underlying cause is thought to be de novo. However having more than one affected sib and no affected parents is usually suggestive of an autosomal recessive disorder or an X-linked one. An exception to this could be variance in penetrance of the disorder, more than one gene or variant being causative and hence dosage effect leading to phenotype (which could explain variance in ID severity level). The final possibility is that a parent could be mosaic to a dominant variation that would not lead in phenotype in his/her case.

From the genomic view, the number of gains and losses is large per sample and narrowing by shared regions allows us to focusing on specific regions under the assumption that there is a heritable region of loss or gain that is directly the cause of mental retardation. In the case that there is an inherited mutation that affects genes involved in maintaining DNA integrity and are involved in replication, a variation could be producing de novo mutations that would cause mental retardation. Hence the cause could be indirect and ID causing CNVs might not be shared across siblings but rather the original mutation that leads to disruptions in DNA integrity.

Results from the Affymetrix cytogenetic array can be viewed and grouped in several ways In this study, we focused on the Toronto Database of Genomic Variation to test if the variation was previously reported in healthy individuals. This does not completely dismiss the fact that it can still be causative of ID as the mental state of the sampled individuals in the reported study might not have been looked at. Also, the reported regions could have a population specific association with mental retardation that has not been investigated before.

This method was initially proposed for use in the Kuwait Medical Genetic Centre. However, the Centre has no link to local research facilities and while the machines were used in the shared research facilities of Kuwait University, the results require a great deal of research and further confirmatory testing that is not possible within a diagnostic laboratory. Technical support is also required in laboratory to set up and use the software analysis. Interpreting the results is challenging, and may not be possible for a clinician, who uses the software infrequently, to interpret results correctly. This testing procedure is therefore not appropriate for a small genetics centre to run in-house with limited research support. It might be feasible on an international collaborative basis, with stringent quality control measures set to ensure reliability and reproducibility of results, and confirmatory testing procedures available to lead to diagnosis that would benefit the family. As a next step, however, efforts are made to establish a bioinformatics unit that would help with building a database, maintaining and labelling data for easy comparison across different laboratory techniques to support in data analysis.

In summary, using the cytogenetic arrays allowed the viewing of several micro gains and deletions and to have a look on a genomic level with a higher resolution than classical karyotyping. It is also be a powerful molecular technique for detecting homozygous regions, which are common in the Kuwaiti population. The results still need further investigation before a proper diagnosis can be reached and a collaborative effort of a team of experts. The Centre is taking steps to establish better analysis support and a pipeline for embracing new technology.

Chapter 7- Discussion

The aim of this thesis was to evaluate different laboratory techniques to detect aberrations, namely copy number variations (CNVs), which could be the underlying genetic cause of ID in the selected families. Based on the results, recommendations were made for establishing a clinical genetic protocol for ID in Kuwait.

Moreover, the methods used in this thesis include Affymetrix cytogenetic arrays 2.7M, MLPA kits, and direct sequencing using ABI sequencer, details for these methods are included in chapter 3. In addition, two samples were sent off for exome sequencing. These methods were used to detect genetic aberrations in DNA samples from two sample sets, one from family members with ID in 16 families and the second from cases affected with RTS.

The study was limited in the number of samples tested and even more so in the reagents supplied. A few genes of interest were identified for further research, which could include reruns, confirmatory testing with other methods, and testing other family members and relatives when possible.

It is worth emphasizing that identifying the underlying genetic cause for ID in each family and setting a protocol to deal with such cases is vital for further support in genetic counselling for family planning purposes and even premarital counselling in these large extended families where consanguinity is still a common practice.

In this chapter, the results and use of each method will be evaluated and discussed with recommendations provided to improve the diagnostic service on offer at Kuwait Medical Genetic Centre workflow, aberration detection, laboratory set up, and follow up on the families tested.

7.1 Previous studies of ID in Kuwait

Internationally, ID is widely studied with research programmes ongoing in all aspects of the disorder. Matson and Boisjoli (Matson and Boisjoli 2009) produced an overview of the research published on intellectual disabilities, describing ID as the "most complex and frequently studied", with around 2000 papers being published per year and genetics being the most popular topic. In comparison, few papers are produced from the Arab world. In the last 10 years only two papers were published on the Kuwaiti population with ID related disorders: on Fragile X syndrome (Bastaki, Hegazy et al. 2004) and on Autism spectrum disorders (Faras, Al Ateeqi et al. 2010), although around 500 cases are referred per year for testing to the Kuwaiti Genetics Centre (as per the registry in the Genetic Centre) based on low IQ. Identifying cases usually occurs after the child fails at certain developmental milestones and not through dysmorphic features, except for some ID related disorders.

This highlights that research in ID, in this part of the world, is very limited and almost non-existent. There is no set policy or vision in either studying these cases further or providing answers to the underlying genetic causes. This deficit must be addressed as advances in research and technology could provide answers to families and the appropriate genetic counselling for family planning purposes or even premarital counselling.

As discussed in chapter 1, the current procedure for identifying ID causing aberrations in Kuwait is limited to certain conditions and further research not being considered of high priority, unless the parents are opting for IVF (in-vitro fertilization) in which an answer is required to proceed with the treatment. Research on ID is therefore of limited scope, although many family members can be affected especially where cousin marriages are allowed. Unless the causative aberration is known, there is no test that can be offered to assist the genetic counsellors with family counselling sessions.

The research infrastructure in Kuwait is inadequate, but laboratories abroad can assist in this research. However, the strict regulations in sending DNA samples abroad must be relaxed to allow for better collaborations. The Ministry of Higher Education and Ministry of Health have invested in scholarships such as the one funding this project. Having the projects focused on research that will benefit the health system should be made a priority with specific relevant topics pre- identified to meet the needs of the health system.

A further area of concern is that little is known about the distribution of genetic variants in the Kuwaiti population. This is being addressed with Kuwait joining the Human Variome Project, by establishing the country node. The aims, as stated in the Human Variome Project plan in the third meeting "Project Roadmap 2012-2016 (Kohonen-Corish, Smith et al. 2013) are to "enables the routine collection, curation, interpretation, and sharing of information on all human genetic variation." This involves setting laboratory standards, ethical practice, knowledge sharing, and capacity building. This is an important step for collaborative efforts in data sharing for comparison of reported variations in affected individual.

The research in this thesis has focused on two clinical areas: the idiopathic form of intellectual disability (ID used throughout), and RTS. RTS syndrome was selected for inclusion in this study as it is a distinct syndrome, and although rare, the underlying cause could shed light on further understanding of biological pathways and genes involved, as ID is one of the clinical features found in RTS. Mutations in the CREBBP gene are found in, 60% of RTS cases, and CREBBP has a role in memory (Martin and Sun 2004; Sharma, Jadhav et al. 2010; Barrett, Malvaez et al. 2011), which is an important aspect of intelligence and is measured within the IQ score.

7.2 Family selections and sample collection

This covers the clinical genetic service provided at the Genetic Centre, including; patient registry, genetic counselling service, medical records, data retrieval, patient follow up, sample collection, storage and handling.

7.2.1 Patient registry

The patient registry at the Kuwait Medical Genetic Centre (KMGC) was the first point for sample selection. Pre-requisites were applied to select families to include in the study. The lack of an electronic archival system meant long hours spent going through hard copy files through the registry. The file numbers were selected based on DNA availability where an electronic system is set for DNA banking (of all the DNA samples at the KMGC) and the syndrome is indicated. In addition there were a number of families included as recommended by the clinical geneticist based on the selection criteria of this study, in which case the family was contacted for consent and blood sample.

As common with other developing countries, Kuwait still needs to unify its effort in setting a national registry to facilitate research studies, especially for complex disorders. Moreover, there is a need for electronic medical records that can ease the search process, particularly, if a new laboratory procedure was made available and patients need to be re-contacted. Also, having the files backed up will reduce the possibility of missing files, as was the case for 2 families that were excluded from the study because the files could not be located.

Electronic registration of the DNA samples has been started, however, some data are still missing and the process is not standardised for all cases that visit the centre, nor does it encompass the whole medical history and a backup system. This means a lack of real statistical data that will enable calculations of incidence and prevalence of

specific disorders. In addition, there are difficulties in having access to the targeted groups for study purposes.

Establishing a specialized database, especially for those disorders that will require further research and would benefit multiple individuals and families, is highly recommended.

Also, having a smooth written process that will deal with sample storage, privacy of information (taking consent where consent is needed) and easier access to target samples would further enhance future genetic analysis. It is worth noting that the Genetic Centre is not a research facility and hence storage was made in order to follow up on diagnosis and for future use. On the other hand, for certain disorders, research is required for diagnosis and the Genetic Centre must establish good relationships with research institutes in order to offer better services for patients. Kuwait must invest in a national strategy for genetic research that would serve the diagnostic services that are offered as part of the public health system. An important part of this is proper storage of information and the retrieval process.

7.2.2 Genetic counselling

Genetic counsellors rely heavily on information to guide their counselling sessions, this includes the family history that is revealed and known causes to the syndrome in question, that includes specific genetic aberrations that can be tested.

It is worth noting that some of the families do not consent for test results to be shared due to fears of social stigma that might impact things such as marriage proposals to healthy sibs. Traditionally, some ID patients were kept away from social circles so that they would not impact marriages to the healthy sibs. This has now reduced with awareness and more and more ID patients enjoy full social lives and a level of acceptance. A similar awareness towards the importance of sharing information and

including relatives in studies to ease the understanding of the underlying cause is important. Confidentiality is respected, however, families must be made aware of the importance of including some family members for research purposes.

In line with creating awareness about the need for research involvement, there is also a need to address the impact consanguinity has on genetic disorders, especially for those with a positive family history. Without knowing the underlying cause it is difficult to advise on the probability of having a child with ID when a family history is present. Likewise, knowing the cause does not guarantee a healthy baby. These aspects of family planning and premarital counselling must be addressed through a large media campaign targeting high school and university students, who are at the marrying age in Kuwait.

7.3 Materials and methods

The laboratory methods selected for this study were primarily based on machine availability and access to laboratories in Kuwait. This includes the Affymetrix machines, which were available for array analysis at the core research facility of Kuwait University, and were considered for purchase for use in diagnostics at the Genetic Centre. Likewise, the MLPA kits and sequencing kits were both run on ABI sequencers available at Kuwait University and the Genetic Centre, more details can be found in chapter 3.

7.3.1 Reagent delivery and technical support

There are practical challenges to performing genetic research in Kuwait, where standard reagents and molecular equipment is not routinely available. All reagents are

manufactured abroad and are imported through various local suppliers and agencies, often lacking technical support.

For example, delivery restrictions delayed progress of the research for months, with the MLPA kits specifically taking over one year to arrive. For efficient workflow and better diagnostic service, this shortcoming of the delivery service must be fixed. It was easier to get certain reagents from abroad through courier mail, however, for temperature sensitive reagents this would be more difficult.

Delays were faced from placing orders, to tracking orders and final drop off. This must be optimized as often some items have short shelf life and so will need to be ordered for specific studies when required. If these are kept for half of their shelf life on storage shelves in warehouses before making their way to the laboratory then their usage will be limited.

In addition, shortage of one reagent meant even more delays as kits would not be used without that missing reagents and local suppliers usually place their orders upon request and do not keep reagents in local stores.

With the lack of technical support when it comes to laboratory setting and data analysis software, the laboratory process would not be run efficiently and might impact quality assurance. These methods are often sensitive and rely on an optimization process to achieve best quality of results, hence the need for technical support or training to ensure better performance.

7.3.2 DNA extraction method

The methods used for DNA extractions vary in yield output (details of laboratory methods can be found in chapter 3); with phenol/chloroform method having higher yield, although the Maxwell run extraction can be programmed for a higher yield

method. This is not often the case in the Genetic Centre as the DNA is extracted for specific testing and large yield is not targeted, on the other hand high yield is vital for research. Also, the phenol in one method affects certain laboratory techniques such as sequencing and array hybridization. It is important to record the method of extraction, as some laboratory techniques are sensitive to it and even sensitive to samples being compared while they have a different extraction method. This is the case for MLPA where variations in extraction method would affect the results of this sensitive test. MLPA results are relative across samples and variants could be reported due to the factor of using different extraction method rather than a true copy number variation. Hence, having one method of DNA extraction or recording the method with the sample data would benefit future research. Great care should be taken at this stage, as DNA is the primary material for most of the laboratory techniques carried out in the centre.

An optimal DNA extraction method that suits the methods in question should be identified for better results. This can be done by comparing the DNA concentration and number of failed samples based on method of extraction.

7.3.3 General laboratory set up

Two locations were used for the laboratory work. One location used laboratories in separate buildings, which created a safety hazard in terms of having to carry sample plates with gloves and no special setting for such procedure. If machines used for one technique were kept in one building that would be better, alternatively access can be made so that samples can be transferred through special access and not be hazardous.

7.4 Cytogenetic arrays

Cytogenetic arrays have the advantage of giving insight into variation on a genomic level. The cytogenetic arrays selected were new in the market at the time and although the software and kits are heavily marked as being for research purposes only, local agents are marketing them as a diagnostic method for routine service.

Nevertheless, many laboratories recommend array technology as a first-tier test for ID and rare disorders with unknown genetic causes (Gijsbers, Lew et al. 2009; Miller, Adam et al. 2010; Vermeesch, Brady et al. 2012). The authors suggest that array technology should replace G-banded karyotyping, which should be reserved for cases suspected of a known chromosomal syndrome. This is partly due to the higher diagnostic yield observed (Miller, Adam et al. 2010) with 15 to 20% yield produced by arrays compared to 3% through G-banded karyotyping (not including major known chromosomal syndromes). This higher yield is mainly due to the higher resolution offered through array analysis, although it does not cover the whole genome, it is able to detect small aberrations. However, some cases are better analysed through G-banding, which can detect balanced translocations and these include cases with known family history or multiple miscarriages.

The challenge of using arrays lies in determining the clinical significance of the variants identified. These aberrations of deletions and duplications can be classified as a) pathogenic or clinically relevant, b) clinically significant but unrelated to phenotype, c) variant of uncertain clinical significance (VOUS), d) benign or polymorphic CNV. Several laboratories have different criteria for each classification with some keeping internal databases for classification in comparison to known pathogenic CNVs or known benign ones. There is a growing consensus of the need to standardise the classification method, with recommendations published by the American College of Medical Genetics and the international system for human cytogenetic nomenclature (ISCN) to

ease the process of data sharing and accurate comparison (Brothman, Persons et al. 2009; Kearney, South et al. 2011; Kearney, Thorland et al. 2011).

There were several limitations for the use of this technique in this study. First, evaluation of this method would have been better if a positive sample (i.e. one with a known aberration) was used to assess the accuracy of the aberration detection of this method. Qiao and colleagues for their study carried this out before detecting aberrations in new cases (Qiao, Tyson et al. 2013). Vermeesch recommends 50 samples as being a reasonable size for validation of normal and known abnormal samples (Vermeesch, Brady et al. 2012). Second, the reference sample was based on the one supplied by the software and not one that would incorporate ethnic variation. Third, the experimental design affected analysis. Having more members of the family enables the narrowing down of results, with a recommendation of testing trios (parents and affected case) as an optimal design, since testing the parents is required for determining the origin of the variations and determine possible causation. Fourth, there were no supported confirmatory methods in the laboratory, e.g. MLPA could be considered a secondary confirmatory method but it needs to match the covered CNV site. There should be an established process for confirmatory testing.

The arrays are supplied with freely available Chromosomal Suite software but there are several other software packages available for analysis. Comparing results with software such as Nexus and PennCNV would have been useful in adding more confidence where findings are complementary. This was not possible for this study because Nexus software required expensive licensing and the other extensive training for analysis. This raises the point of the need to have informatics assistance, as the results are not as straightforward as they are marketed.

Moreover, the arrays could be useful in identifying major aberrations or if a specific CNV known to associate to syndrome was suspected. However, when the cause is unknown and heterogeneous such as the case of non-specific ID then the analysis is

not as straight forward. Different stringencies and levels of data analysis would reveal different results and require further research.

It is vital to link up with research centres for this analysis, in order to identify the causative aberration and confirm with a secondary confirmatory test to make sure that this is the cause of the disorder. This workflow must be established i.e. testing arrays then secondary testing, followed by testing the parents and then other family members, when possible.

A key issue in the experimental design was having more family members, i.e. another affected sib or unaffected parent, which made it easier to select potentially pathogenic CNVs. An ideal design would use family trios. Also, comparing the results to the Cooper list of potentially pathogenic CNVs (Cooper, Coe et al. 2011) was of great assistance in selecting some CNVs. Other such lists exist and data sharing between laboratories and establishing internal databases would benefit diagnosis significantly.

Another issue was the complications of getting genotype data and extracting SNP marker information from the chromosomal suite software for further analysis with other software. This was discussed with Affymetrix and they sent an agent from Abu Dhabi to deal with the issue and it was explained that the arrays used were not widely marketed and that initially there were restrictions on genotype data extraction to protect identifying sample and for anonymity. Genotype data are usually exported from another software, genotype console, which requires the raw data in another file format that was not collected and saved at the time. These files were not kept, as the Chromosomal suite software used requires the CEL file format. However, in the new version of the software this information is available within the Chromosomal suite software and genotyping data is available for viewing and exporting to other software. Unfortunately, the new version of the software does not accept the CEL file format and is relying on the on raw data format to be used to extract and analyse all data.

7.5 MLPA kits

The MLPA kits were easy to handle in the wet laboratory and the results easy to follow once the analysis software was set and carried out. Limitations would be in the reference sample used, as these were limited and some variants were observed in one of the control samples used for reference. Validation with a sample with known CNV corresponding with those detected by the kit would add to the confidence and reliability of the results. Variations reported could be caused by other factors such as the DNA extraction method or sample storage, which would affect the sensitivity and hence the results, as it measures relative difference. The method was easy to handle in comparison to arrays but limited in coverage with one exon per probe at a total of 40 probes. However, for analysis of a large number of samples it could be useful when certain mutations are suspected.

7.6 Direct sequencing

Although well established, the results generated from direct sequencing of the CREBBP gene required further reruns to confirm findings just to ensure that the variant is not a false signal or due to wrong base calling, as in some of the cases there was no corresponding peak to the base. Confirmatory analysis was not possible as part of this research due to lack of availability of reagents. In addition to running RTS samples, the original aim was to sequence the CREBBP gene on the ID and normal samples but this was not possible due to problems with reagents, which were limited and took a long time to order. In the RTS samples, sequencing was limited to specific exons, and hence a mutation could still lie outside of these exons or within intronic regions not covered by the primers. Having a perfectly aligned sequence in exons 18 to 30 does not mean that the CREBBP gene is normal, as the other exons were not tested.

7.7 Overall variation results

The results of this study were inconclusive as further confirmatory testing is required for all the methods used. Nevertheless, there were a number of regions of interest. Main families of interest to study further are family 7 and A.

7.8 Further testing of families

Overall, potentially pathogenic CNVs found from the arrays should be tested with another method, followed by testing the parents for the same CNV. If multiple CNVs are suspected the parental samples can be run on arrays before narrowing down further and then testing relatives.

For the families that failed to give any results (families 12 to16, and C, E, F, and H), blood samples must be requested again as failure could be due to the quality of DNA.

For other families, the suspected pathogenic variants should be followed up before going back to the array data and investigating the other CNVs. The assumed underlying model of inheritance for ID in this study is an autosomal recessive one that is shared amongst sibs. If variations are de novo due to a shared impairments in the transcription process then these variants could be missed when selecting based on shared regions.

MLPA should be optimized and rerun to include more normal samples for reference.

For family 1, two shared CNVs of interest were shared, one spanning 42 genes on chromosome 8 and judging by the number of genes it could be significant. Confirmation via another method such as FISH and then testing the parents would be

the next steps. Searching for the function of each gene separately might give clues on potential functional role.

For family 2, two shared CNVs of gain on chromosome 8 were observed and should be followed up and confirmed by another method before testing the parents. The same CNV was also found for samples in families 3, 6, 7, and 10, it could be a normal variant, as indicated by the high number of studies on the DGV, however, this region covers many genes and perhaps some of these have not been extensively studied.

For family 5, another family member can be tested with the array to locate shared regions. For families with no potentially pathogenic CNVs identified in the different selection methods, the selection criteria must be adjusted or the list of genes can be compared to other databases. Running normal samples on arrays would be useful for compiling potentially benign CNVs.

Family 7 gave the most interesting results based on the experimental design, in which two affected sibs and a normal parent were tested on the arrays and then DNA from the two affected sibs were sent to exome sequencing, resulting in two potentially pathogenic single nucleotide variations (SNVs). This can be followed via direct sequencing and cosegregation with phenotype should be tested within this family.

For family 9, SEPT2 gene can be followed that has been reported in the Cooper list. , The CNV was also observed in the unaffected father, but might not be fully penetrant. This region can be investigated further or DNA reordered as the array results of both family members had many variations detected.

For family 10, several genes matched those reported in the Cooper list but the copy number type was different and hence further investigation into databases or literature would be recommended to find out if CNVs of gains are also pathogenic. These can then be followed further.

For families 3, 11, D and J further runs to confirm MLPA results will be required and to see if these cosegregate with the phenotype.

For family A, there were three CNVs of interest covering many genes that are shared amongst the affected sibs that can be investigated further. In addition, one CNV of gain matched covered a gene reported as pathogenic when in a gain status. This would be the first point of investigation. Also a gain encompassing gene EP300 in one of the sibs can be investigated further.

For the RTS families, direct sequencing should be repeated for suspected variation and parents should be tested to establish that it is a de novo variation. In any case a mutation in CREBBP gene would only account for 60% of cases and in this study focus is on around 10 exons of the total 30 exons. Placing samples on arrays and comparing the samples together could identify possible other genes involved in the pathogenicity of RTS.

For arrays, results of mosaicism were available but not investigated further; these could be looked into if no other CNVs were found to be causative.

7.7 Recommendations and future work:

7.7.1 Investment in research

Kuwait would benefit in setting a strategy for genetic research and perhaps by starting with international collaborations that will bring benefit from experts in the field and standardizing laboratory procedures being used before investing in different machines.

A research strategy should incorporate training of staff and sending them for various training and scholarships abroad, as it is vital to set a highly competent team when dealing with research and data interpretation. Hence, as an outcome to this research

workflow, laboratory standards can be set based on the selected methods and aberrations identified. Learning from best practice in laboratories with high international standards would ensure better quality assurance.

International collaboration and joining projects such as the Human Variome project should be encouraged and supported to benefit from other countries' findings and research, with similar studies being conducted in Saudi, collaboration will be beneficial to the ethnically similar nations.

Furthermore, there are a number of rare disorders that are found in Kuwait such as RTS and partly due to consanguineous marriages. These should be included in gene detection research as this could further our knowledge and understanding of important biological pathways.

7.7.2 Establish a national registry and electronic medical record

This can be useful for syndromes that need further studies such as ID as well as rare disorders. A well-organized electronic medical record that also secures the privacy of patient information should support this. This well kept database can be used for future research and having a DNA bank where individuals consent to further studies, would benefit the nation as a whole and families in the future. Internal databases are important for array studies in particular.

7.7.3 Raising awareness

For further participation in research and giving consent the public must be aware of the value of research and impact of informed decisions when it comes to family planning or premarital counselling. Encouraging participation and breaking away from

social stigma could unite efforts to further understand syndromes such as ID and even perhaps open doors for research into intervention strategies in the future.

In addition, the impact of consanguinity on the chances of having a child with a genetic disorder must be communicated to youth in order for them to make an informed decision, especially in cases where there is a family history of syndromes such as ID.

7.7.4 Reagent supply and technical support

The Ministry of Health must review the order and delivery policies and procedures to ensure constant supply of vital reagents on time. Setting standard orders required for each method and ensuring that suppliers provide the necessary training or be able to assist in trouble shooting is essential for smooth operation.

7.7.5 Optimization of laboratory procedures and standards

The laboratory should employ best practice based on international standards. Several certifying bodies exist to ensure the laboratory is set to high standards and the quality of testing can be assessed for quality assurance and reliability. Good practice includes recording the methods such as the type of DNA extraction method, and applying a labeling format as part of a standard procedure. This has been implemented recently due to issues raised during the work of this project.

For cytogenetic arrays several measures can be made: 1) Validation of detection by running samples positive samples with known aberrations and known negative (normal with no family history) to test for the accuracy of detection that the procedure has been run correctly and at a right optimization. 2) Creating local reference files, for comparison that might take into account ethnic variations, 3) Run duplicates as

another internal validation step to measure reproducibility, 4) Optimize analysis settings by defining optimal calling threshold based on the results from normal samples. 5) Set a workflow for confirmatory testing, several have been suggested (Gijsbers, Lew et al. 2009; Miller, Adam et al. 2010; Vermeesch, Brady et al. 2012), this depends on the class of CNV. For potentially pathogenic CNV the parents should be tested as well with one method such as FISH, MLPA, direct sequencing, or even exome sequencing making sure the CNV of interest is covered, and if the CNV is found in the parents then it will be considered potentially benign. Nevertheless, incomplete penetrance could be involved. Cosegregation of variant with phenotype is important to establish potential pathogenicity and having more family members to test for that would increase the power and reliability of the results.

For MLPA set up, having positive controls will benefit validation efforts and add confidence to the accuracy of the results. Each kit should be optimized accordingly before use.

Moreover, for direct sequencing, supplies should be available for reruns and running ethnically matched normal samples would be important especially if a repeated variation is observed.

7.7.6 Applying a population screening and family testing work flows

Comparing across methods two workflows are suggested one (workflow 1) for population screening for known CNVs and the other (workflow 2) for detecting novel pathogenic CNVs within families.

Workflow 1, as illustrated in figure 52 below, is ideal for establishing the CNV frequency of specific CNVs that have been linked to ID or related disorders. Since this has not been previously done for the samples in the Kuwait Medical Genetic Centre then this can be done using the MLPA kits P245, P036 and P070. This can include

samples from patients with no previous family history. Different kits can be used and abnormality confirmed with another method such as FISH. Samples with no abnormality detected can then be moved to workflow 2, where more expensive methods can be used to identify abnormality through CNV detection or homozygosity mapping technique as illustrated in figure 53 below. If one family is investigated then workflow 2 would be the first choice.

7.7.7 Developing informatics support

The data analysis of array results needs further support and results might not be possible to interpret by the laboratory technician or the clinician. If this support is not established locally at least ties should be made with international centres to provide such services, otherwise the results will be difficult to interpret for second stage confirmation and further testing. A clear process should be established for standard workflow and to identify the confirmatory methods to be used for compatibility purposes. For instance MLPA covers specific exons that might not match those identified via the arrays and so another method is required for confirmatory testing.

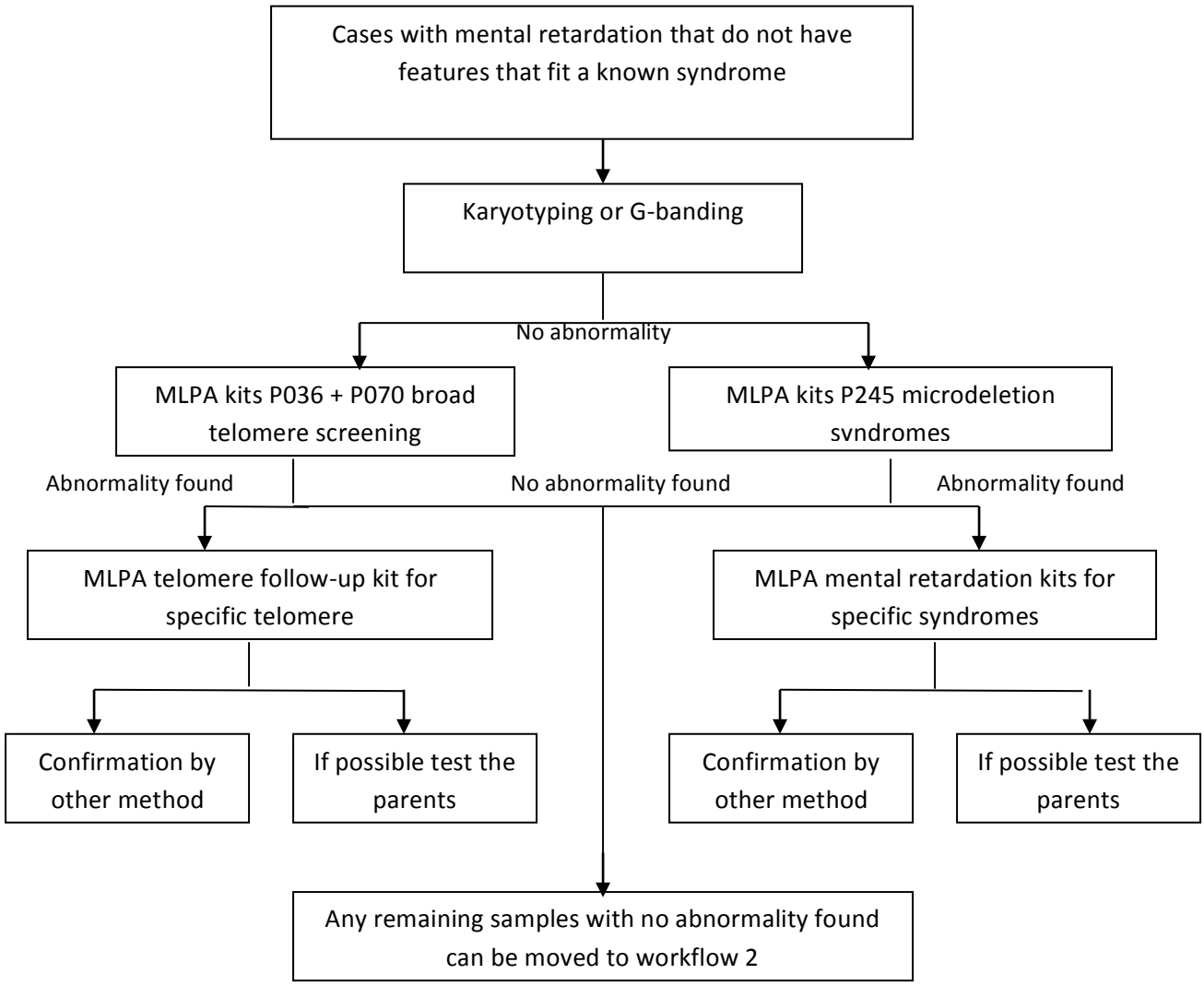
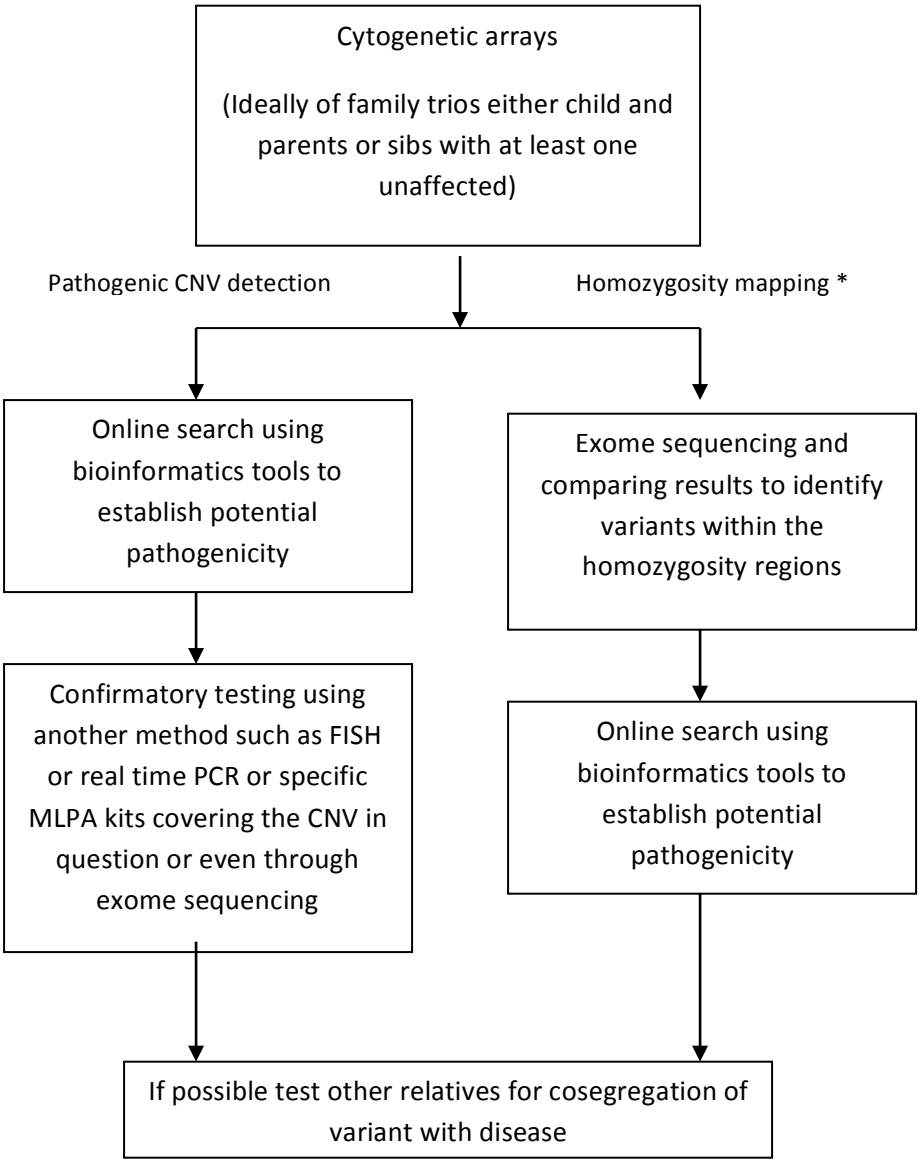


Figure 55 – WorkFlow 1 for population screening or large number of samples



* (Only relevant if an autosomal recessive mutation is suspected)

Figure 56 WorkFlow 2 for family testing with unknown causative gene

Moreover, data storage should be addressed especially in the case of the arrays where large data sets were generated and can be stored for future studies; this internal database can be supported by different software (Vermeesch, Brady et al. 2012).

This need was communicated with the Genetic Centre leading to the plan of establishing of a specialist bioinformatics unit, lead by the author, that will support result analysis including training staff on basic tools, data storage and database design creating scripts for specific searches. At this point assessment of needs is being drafted based on: the laboratory techniques carried out by the centre, results file formats, reports required, and new techniques to be implemented.

7.7.8 Other genetic causes

The laboratory methods used here were very specific in identifying certain regions of variations at a DNA level. There are other genetic causes that include variations on splicing that would affect the RNA. Other epigenetic variations such as methylation methods have been attributed to several genetic syndromes and would not be detected using these methods. These should be explored for future testing when no variation is confirmed to be pathogenic.

7.7.9 Ethical consideration

Several ethical issues are raised, especially when dealing with the arrays that must be addressed pretesting. This would be with regards to clinically significant CNVs but not related to the initial referral syndrome and are considered mainly as susceptibility or predisposition loci. Since the array tests millions of probes at a time it opens a window into information for other disorders and susceptibilities. If the family in question wants to know this information it must be stated before the test is done and reported separately. However, it must be communicated properly that susceptibility does not imply that the disease will be certain e.g. in cases of cancer or heart disease.

Another issue is the testing of minors, who should be dealt with great consideration especially when looking at susceptibility loci that they might not want others to know about and is not part of the syndrome in question.

Furthermore, when testing other family members the information should be kept confidential when consent is not given to reveal results. However, this may create pressure from relatives to know especially if they are being involved in the study. These issues must be raised for consideration.

In addition, current understanding of the significance of the results must be properly communicated in counselling sessions. These are preliminary results and will be subject to further testing and unless a CNV is confirmed to be pathogenic then it should not be the basis of a decision that family members should be making such as opting for an IVF treatment or deciding on getting married. Likelihood and clinical significance must be clarified for proper decision-making.

7.7.10 Limitations of the study

A specific challenge in this research was to select samples with a similar level of ID, since using patients with similar phenotype could perhaps identify a common cause. This was not possible for some families where different individuals had different levels of ID, hence, a possible variance in penetrance may be suspected if the same genetic cause is assumed. It was the assumption that the genetic aberration was family specific but for further association studies perhaps the ID level would be significant in family selection. Having this information available as well as the different components of the test, e.g. memory score could be useful for further research.

In addition, having positive samples is vital for validating methods especially when they are tested for the first time. As well as setting reference samples that are ethnically matched as CNVs and SNPs are population specific and there could be common

polymorphism not found in other ethnic groups, especially that the Kuwaiti population is not extensively studied.

The major limitation of the study was the limited availability of reagents, which restricted the number of runs and samples placed on the array. Having more samples from family members on arrays was more informative.

7.7.11 Embracing new technology and building collaboration

New technology is constantly being introduced into the market with the Affymetrix cytogenetic arrays used in this study being discontinued and replace with new CytoScan HD arrays with a new version of the software that allows genotype data to be displayed using the Chromosomal Suite software features. Also, specific arrays for intellectual disability, called CytoScan Dx Assay, have been developed by Affymetrix, as the first tier aid in diagnostics of ID. These arrays are compatible with the machines and washing stations used in this study.

In addition, new Hiseq and Miseq next generation sequencing system by Illumina has been installed in a research laboratory, Dasman Diabetes Institute, in Kuwait and the Genetic Centre has set a pilot collaboration on a number of unsolved cases to identify the underlying genetic cause. A plan will be made to utilise this technology and investigate RTS families first to identify the genes involved. It is vital the Centre links to a research facility as there are many unsolved cases that require further investigation.

In summary, the testing of these families using the different methods was the first to be performed in Kuwait. A proper workflow is required were methods are optimized and validated to offer better services. Moreover, testing with arrays can be a useful first tier technique for ID but more family members should be tested to narrow down the potentially pathogenic CNVs. This method will also benefit from curating a local database of variations. Another confirmatory testing must be optimised that can vary

between MLPA, FISH or even real time PCR. Also, sending samples off to exome sequencing could provide useful insight on potential pathogenic SNVs harboured in regions of homozygosity. Hence, easing the process of sending samples abroad could benefit and improve the diagnostic services on offer for these families. A strategy should be prepared to set collaboration with research centres and utilising new technology to utilise advancements in science.

Identifying the causative genetic factor in these families for ID and even rare disorders such as RTS would benefit clinicians in gaining information vital for a better genetic counselling session for better informed decisions.

References

Abou Jamra, R., S. Wohlfart, et al. (2011). "Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity." Eur J Hum Genet **19**(11): 1161-1166.

Ahn, J. W., C. M. Ogilvie, et al. (2007). "Detection of subtelomere imbalance using MLPA: validation, development of an analysis protocol, and application in a diagnostic centre." BMC Med Genet **8**: 9.

Al-Kandari, Y. Y. and D. E. Crews (2011). "The effect of consanguinity on congenital disabilities in the Kuwaiti population." J Biosoc Sci **43**(1): 65-73.

Al-Owain, M., H. Al-Zaidan, et al. (2012). "Map of autosomal recessive genetic disorders in Saudi Arabia: concepts and future directions." Am J Med Genet A **158A**(10): 2629-2640.

Alkuraya, F. S. (2010). "Homozygosity mapping: one more tool in the clinical geneticist's toolbox." Genet Med **12**(4): 236-239.

Alkuraya, F. S. (2013). "The application of next-generation sequencing in the autozygosity mapping of human recessive diseases." Hum Genet **132**(11): 1197-1211.

Arlt, M. F., T. E. Wilson, et al. (2012). "Replication stress and mechanisms of CNV formation." Curr Opin Genet Dev **22**(3): 204-210.

Barrett, R. M., M. Malvaez, et al. (2011). "Hippocampal focal knockout of CBP affects specific histone modifications, long-term potentiation, and long-term memory." Neuropsychopharmacology **36**(8): 1545-1556.

Bartholdi, D., J. H. Roelfsema, et al. (2007). "Genetic heterogeneity in Rubinstein-Taybi syndrome: delineation of the phenotype of the first patients carrying mutations in EP300." J Med Genet **44**(5): 327-333.

Bartsch, O., W. Kress, et al. (2010). "Inheritance and variable expression in Rubinstein-Taybi syndrome." Am J Med Genet A **152A**(9): 2254-2261.

Bartsch, O., K. Locher, et al. (2002). "Molecular studies in 10 cases of Rubinstein-Taybi syndrome, including a mild variant showing a missense mutation in codon 1175 of CREBBP." J Med Genet **39**(7): 496-501.

Bartsch, O., S. Rasi, et al. (2006). "Evidence for a new contiguous gene syndrome, the chromosome 16p13.3 deletion syndrome alias severe Rubinstein-Taybi syndrome." Hum Genet **120**(2): 179-186.

Bartsch, O., S. Schmidt, et al. (2005). "DNA sequencing of CREBBP demonstrates mutations in 56% of patients with Rubinstein-Taybi syndrome (RSTS) and in another patient with incomplete RSTS." Hum Genet **117**(5): 485-493.

Bastaki, L. A., F. Hegazy, et al. (2004). "Fragile X syndrome: a clinico-genetic study of mentally retarded patients in Kuwait." East Mediterr Health J **10**(1-2): 116-124.

Bentivegna, A., D. Milani, et al. (2006). "Rubinstein-Taybi Syndrome: spectrum of CREBBP mutations in Italian patients." BMC Med Genet **7**: 77.

Berry, R., A. C. Smith, et al. (1987). "Cytogenetic studies in patients with multiple anomalies with or without mental retardation." Birth Defects Orig Artic Ser **23**(6): 100-110.

Bittles, A. H. (2008). "A community genetics perspective on consanguineous marriage." Community Genet **11**(6): 324-330.
Brothman, A. R., D. L. Persons, et al. (2009). "Nomenclature evolution: Changes in the ISCN from the 2005 to the 2009 edition." Cytogenet Genome Res **127**(1): 1-4.
Chiang, P. W., N. C. Lee, et al. (2009). "Somatic and germ-line mosaicism in Rubinstein-Taybi syndrome." Am J Med Genet A **149A**(7): 1463-1467.
Christofolini, D. M., M. A. de Paula Ramos, et al. (2010). "Subtelomeric rearrangements and copy number variations in people with intellectual disabilities." J Intellect Disabil Res **54**(10): 938-942.
Cooper, G. M., B. P. Coe, et al. (2011). "A copy number variation morbidity map of developmental delay." Nat Genet **43**(9): 838-846.
Coupry, I., C. Roudaut, et al. (2002). "Molecular analysis of the CBP gene in 60 patients with Rubinstein-Taybi syndrome." J Med Genet **39**(6): 415-421.
Crotwell, P. L. and H. E. Hoyme (2012). "Advances in whole-genome genetic testing: from chromosomes to microarrays." Curr Probl Pediatr Adolesc Health Care **42**(3): 47-73.
de Ligt, J., P. M. Boone, et al. (2013). "Detection of Clinically Relevant Copy Number Variants with Whole-Exome Sequencing." Hum Mutat.
de Vries, B. B., S. M. White, et al. (2001). "Clinical studies on submicroscopic subtelomeric rearrangements: a checklist." J Med Genet **38**(3): 145-150.
Ellison, J. W., J. A. Rosenfeld, et al. (2012). "Genetic Basis of Intellectual Disability." Annu Rev Med.
Faras, H., N. Al Ateeqi, et al. (2010). "Autism spectrum disorders." Ann Saudi Med **30**(4): 295-300.
Foley, P., D. Bunyan, et al. (2009). "Further case of Rubinstein-Taybi syndrome due to a deletion in EP300." Am J Med Genet A **149A**(5): 997-1000.
Franklin, T. B. and I. M. Mansuy (2011). "The involvement of epigenetic defects in mental retardation." Neurobiol Learn Mem **96**(1): 61-67.
Galasso, C., A. Lo-Castro, et al. (2010). ""Idiopathic" mental retardation and new chromosomal abnormalities." Ital J Pediatr **36**: 17.
Gervasini, C., P. Castronovo, et al. (2007). "High frequency of mosaic CREBBP deletions in Rubinstein-Taybi syndrome patients and mapping of somatic and germ-line breakpoints." Genomics **90**(5): 567-573.
Gervasini, C., F. Mottadelli, et al. (2010). "High frequency of copy number imbalances in Rubinstein-Taybi patients negative to CREBBP mutational analysis." Eur J Hum Genet **18**(7): 768-775.
Ghosh, A. K. and J. Varga (2007). "The transcriptional coactivator and acetyltransferase p300 in fibroblast biology and fibrosis." J Cell Physiol **213**(3): 663-671.
Gijsbers, A. C., J. Y. Lew, et al. (2009). "A new diagnostic workflow for patients with mental retardation and/or multiple congenital abnormalities: test arrays first." Eur J Hum Genet **17**(11): 1394-1402.
Gilissen, C., J. Y. Hehir-Kwa, et al. (2014). "Genome sequencing identifies major causes of severe intellectual disability." Nature.

Girirajan, S., Z. Brkanac, et al. (2011). "Relative burden of large CNVs on a range of neurodevelopmental phenotypes." PLoS Genet **7**(11): e1002334.
Girirajan, S., J. A. Rosenfeld, et al. (2012). "Phenotypic heterogeneity of genomic disorders and rare copy-number variants." N Engl J Med **367**(14): 1321-1331.
Grayton, H. M., C. Fernandes, et al. (2012). "Copy number variations in neurodevelopmental disorders." Prog Neurobiol **99**(1): 81-91.
Gropman, A. L. and M. L. Batshaw (2010). "Epigenetics, copy number variation, and other molecular mechanisms underlying neurodevelopmental disabilities: new insights and diagnostic approaches." J Dev Behav Pediatr **31**(7): 582-591.
Hallam, T. M. and R. Bourtchouladze (2006). "Rubinstein-Taybi syndrome: molecular findings and therapeutic approaches to improve cognitive dysfunction." Cell Mol Life Sci **63**(15): 1725-1735.
Harvey, C. G., S. D. Menon, et al. (2007). "Sequence variants within exon 1 of MECP2 occur in females with mental retardation." Am J Med Genet B Neuropsychiatr Genet **144B**(3): 355-360.
He, M., B. S. Tang, et al. (2014). "Using a combination of whole-exome sequencing and homozygosity mapping to identify a novel mutation of SCARB2." Clin Genet.
Heil, K. M. and C. P. Schaaf (2013). "The genetics of Autism Spectrum Disorders--a guide for clinicians." Curr Psychiatry Rep **15**(1): 334.
Hsiung, S. H. (2004). "Rubinstein-Taybi syndrome (broad thumb-hallux syndrome)." Dermatol Online J **10**(3): 2.
Jehee, F. S., J. T. Takamori, et al. (2011). "Using a combination of MLPA kits to detect chromosomal imbalances in patients with multiple congenital anomalies and mental retardation is a valuable choice for developing countries." Eur J Med Genet **54**(4): e425-432.
Kakar, N., I. Goebel, et al. (2012). "A homozygous splice site mutation in TRAPPC9 causes intellectual disability and microcephaly." Eur J Med Genet **55**(12): 727-731.
Kaufman, L., M. Ayub, et al. (2010). "The genetic basis of non-syndromic intellectual disability: a review." J Neurodev Disord **2**(4): 182-209.
Kearney, H. M., S. T. South, et al. (2011). "American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities." Genet Med **13**(7): 676-679.
Kearney, H. M., E. C. Thorland, et al. (2011). "American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants." Genet Med **13**(7): 680-685.
Kent, W. J. (2002). "BLAT--the BLAST-like alignment tool." Genome Res **12**(4): 656-664.
Kent, W. J., C. W. Sugnet, et al. (2002). "The human genome browser at UCSC." Genome Res **12**(6): 996-1006.

Kohonen-Corish, M. R., T. D. Smith, et al. (2013). "Beyond the genomics blueprint: the 4th Human Variome Project Meeting, UNESCO, Paris, 2012." Genet Med **15**(7): 507-512.
Koolen, D. A., W. M. Nillesen, et al. (2004). "Screening for subtelomeric rearrangements in 210 patients with unexplained mental retardation using multiplex ligation dependent probe amplification (MLPA)." J Med Genet **41**(12): 892-899.
Lam, A. C., S. T. Lam, et al. (2006). "High rate of detection of subtelomeric aberration by using combined MLPA and subtelomeric FISH approach in patients with moderate to severe mental retardation." Clin Biochem **39**(3): 196-202.
Lubs, H. A., R. E. Stevenson, et al. (2012). "Fragile X and X-linked intellectual disability: four decades of discovery." Am J Hum Genet **90**(4): 579-590.
Malhotra, D. and J. Sebat (2012). "CNVs: harbingers of a rare variant revolution in psychiatric genetics." Cell **148**(6): 1223-1241.
Mandal, K., V. R. Boggula, et al. (2009). "Use of Multiplex Ligation-Dependent Probe Amplification (MLPA) in screening of subtelomeric regions in children with idiopathic mental retardation." Indian J Pediatr **76**(10): 1027-1031.
Marangi, G., V. Leuzzi, et al. (2008). "Duplication of the Rubinstein-Taybi region on 16p13.3 is associated with a distinctive phenotype." Am J Med Genet A **146A**(18): 2313-2317.
Martin, K. C. and Y. E. Sun (2004). "To learn better, keep the HAT on." Neuron **42**(6): 879-881.
Matson, J. L. and J. A. Boisjoli (2009). "An overview of developments in research on persons with intellectual disabilities." Res Dev Disabil **30**(3): 587-591.
Mefford, H. C., M. L. Batshaw, et al. (2012). "Genomics, intellectual disability, and autism." N Engl J Med **366**(8): 733-743.
Miller, D. T., M. P. Adam, et al. (2010). "Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies." Am J Hum Genet **86**(5): 749-764.
Monfort, S., C. Orellana, et al. (2006). "Evaluation of MLPA for the detection of cryptic subtelomeric rearrangements." J Lab Clin Med **147**(6): 295-300.
Morrow, E. M. (2010). "Genomic copy number variation in disorders of cognitive development." J Am Acad Child Adolesc Psychiatry **49**(11): 1091-1104.
Musante, L. and H. H. Ropers (2014). "Genetics of recessive cognitive disorders." Trends Genet **30**(1): 32-39.
Northrop, E. L., H. Ren, et al. (2005). "Detection of cryptic subtelomeric chromosome abnormalities and identification of anonymous chromatin using a quantitative multiplex ligation-dependent probe amplification (MLPA) assay." Hum Mutat **26**(5): 477-486.
Oike, Y., N. Takakura, et al. (1999). "Mice homozygous for a truncated form of CREB-binding protein exhibit defects in hematopoiesis and vasculo-angiogenesis." Blood **93**(9): 2771-2779.

Palomares, M., A. Delicado, et al. (2006). "MLPA vs multiprobe FISH: comparison of two methods for the screening of subtelomeric rearrangements in 50 patients with idiopathic mental retardation." Clin Genet **69**(3): 228-233.
Petrij, F., R. H. Giles, et al. (1995). "Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP." Nature **376**(6538): 348-351.
Pohovski, L. M., K. K. Dumic, et al. (2013). "Multiplex ligation-dependent probe amplification workflow for the detection of submicroscopic chromosomal abnormalities in patients with developmental delay/intellectual disability." Mol Cytogenet **6**(1): 7.
Qiao, Y., C. Tyson, et al. (2013). "Clinical application of 2.7M Cytogenetics array for CNV detection in subjects with idiopathic autism and/or intellectual disability." Clin Genet **83**(2): 145-154.
Rauch, A., D. Wieczorek, et al. (2012). "Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study." Lancet **380**(9854): 1674-1682.
Rehman, S., S. M. Baig, et al. (2011). "Autozygosity mapping of a large consanguineous Pakistani family reveals a novel non-syndromic autosomal recessive mental retardation locus on 11p15-tel." Neurogenetics **12**(3): 247-251.
Ren, C. M., Y. Liang, et al. (2013). "Balanced translocation t(3;18)(p13;q22.3) and points mutation in the ZNF407 gene detected in patients with both moderate non-syndromic intellectual disability and autism." Biochim Biophys Acta **1832**(3): 431-438.
Rizzi, T. S., G. Beunders, et al. (2012). "Supporting the generalist genes hypothesis for intellectual ability/disability: the case of SNAP25." Genes Brain Behav **11**(7): 767-771.
Roelfsema, J. H. and D. J. Peters (2007). "Rubinstein-Taybi syndrome: clinical and molecular overview." Expert Rev Mol Med **9**(23): 1-16.
Roelfsema, J. H., S. J. White, et al. (2005). "Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease." Am J Hum Genet **76**(4): 572-580.
Rooms, L., E. Reyniers, et al. (2004). "Subtelomeric deletions detected in patients with idiopathic mental retardation using multiplex ligation-dependent probe amplification (MLPA)." Hum Mutat **23**(1): 17-21.
Rooms, L., E. Reyniers, et al. (2006). "Multiplex ligation-dependent probe amplification to detect subtelomeric rearrangements in routine diagnostics." Clin Genet **69**(1): 58-64.
Ropers, F., E. Derivery, et al. (2011). "Identification of a novel candidate gene for non-syndromic autosomal recessive intellectual disability: the WASH complex member SWIP." Hum Mol Genet **20**(13): 2585-2590.
Ropers, H. H. (2008). "Genetics of intellectual disability." Curr Opin Genet Dev **18**(3): 241-250.
Rujescu, D. and D. A. Collier (2009). "Dissecting the many genetic faces of schizophrenia." Epidemiol Psichiatr Soc **18**(2): 91-95.

Sakamoto, K., K. Karelina, et al. (2011). "CREB: a multifaceted regulator of neuronal plasticity and protection." J Neurochem **116**(1): 1-9.
Schorry, E. K., M. Keddache, et al. (2008). "Genotype-phenotype correlations in Rubinstein-Taybi syndrome." Am J Med Genet A **146A**(19): 2512-2519.
Sharma, N., S. P. Jadhav, et al. (2010). "CREBBP re-arrangements affect protein function and lead to aberrant neuronal differentiation." Differentiation **79**(4-5): 218-231.
Soltani Banavandi, M. J., K. Kahrizi, et al. (2012). "Investigation of genetic causes of intellectual disability in kerman province, South East of iran." Iran Red Crescent Med J **14**(2): 79-85.
Stef, M., D. Simon, et al. (2007). "Spectrum of CREBBP gene dosage anomalies in Rubinstein-Taybi syndrome patients." Eur J Hum Genet **15**(8): 843-847.
Stefansson, H., D. Rujescu, et al. (2008). "Large recurrent microdeletions associated with schizophrenia." Nature **455**(7210): 232-236.
Stegmann, A. P., L. M. Jonker, et al. (2008). "Prospective screening of patients with unexplained mental retardation using subtelomeric MLPA strongly increases the detection rate of cryptic unbalanced chromosomal rearrangements." Eur J Med Genet **51**(2): 93-105.
Stuppia, L., I. Antonucci, et al. (2012). "Use of the MLPA Assay in the Molecular Diagnosis of Gene Copy Number Alterations in Human Genetic Diseases." Int J Mol Sci **13**(3): 3245-3276.
Thienpont, B., J. Breckpot, et al. (2007). "A microduplication of CBP in a patient with mental retardation and a congenital heart defect." Am J Med Genet A **143A**(18): 2160-2164.
Topper, S., C. Ober, et al. (2011). "Exome sequencing and the genetics of intellectual disability." Clin Genet **80**(2): 117-126.
Udaka, T., H. Samejima, et al. (2005). "Comprehensive screening of CREB-binding protein gene mutations among patients with Rubinstein-Taybi syndrome using denaturing high-performance liquid chromatography." Congenit Anom (Kyoto) **45**(4): 125-131.
van Binsbergen, E. (2011). "Origins and breakpoint analyses of copy number variations: up close and personal." Cytogenet Genome Res **135**(3-4): 271-276.
Vermeesch, J. R., P. D. Brady, et al. (2012). "Genome-wide arrays: quality criteria and platforms to be used in routine diagnostics." Hum Mutat **33**(6): 906-915.
Verpelli, C. and C. Sala (2012). "Molecular and synaptic defects in intellectual disability syndromes." Curr Opin Neurobiol **22**(3): 530-536.
Viosca, J., J. P. Lopez-Atalaya, et al. (2010). "Syndromic features and mild cognitive impairment in mice with genetic reduction on p300 activity: Differential contribution of p300 and CBP to Rubinstein-Taybi syndrome etiology." Neurobiol Dis **37**(1): 186-194.
Webber, C., J. Y. Hehir-Kwa, et al. (2009). "Forging links between human mental retardation-associated CNVs and mouse gene knockout models." PLoS Genet **5**(6): e1000531.

Wechsler, D. (2004). The Wechsler intelligence scale for children. London, Pearson Assessment.
Yan, J. B., M. Xu, et al. (2011). "Rapid screening for chromosomal aneuploidies using array-MLPA." BMC Med Genet **12**: 68.
Zimmermann, N., A. M. Acosta, et al. (2007). "Confirmation of EP300 gene mutations as a rare cause of Rubinstein-Taybi syndrome." Eur J Hum Genet **15**(8): 837-842.
Zweier, C., E. K. de Jong, et al. (2009). "CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila." Am J Hum Genet **85**(5): 655-666.

Appendices

Appendix for Chapter 4

Direct sequencing results for each exon consisting of a summary table of the blat alignment result for each sample which includes the alignment score, start and end point, query size (Qsize), percentage identity, chromosome, strand (forward or reverse), chromosomal start and end location and alignment span. Missing data (due to sample failure) were highlighted in green. This is followed by a perfect alignment image for the expected fragment.

Table A.4.1 showing the summary Blat results of PCR products for the RTS samples using exon 18 primer sets

	Exon 18 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 18	365	1	365	365	100.00%	16	-	3807752	3808116	365
18_1	196	11	274	369	87.50%	16	-	3807827	3808092	266
18_2	38	10	48	370	100.00%	16	-	3808052	3808091	40
18_3	329	1	336	360	99.20%	16	-	3807774	3808111	338

18_4	311	6	325	425	99.10%	16	-	3807780	3808103	324
18_5										
18_6	325	6	341	359	98.60%	16	-	3807780	3808117	338
18_7										
18_8										
18_9	322	6	337	497	98.80%	16	-	3807780	3808115	336
18_10	318	3	330	358	98.80%	16	-	3807778	3808109	332
18_11	329	2	380	416	98.60%	16	-	3807778	3808331	554
18_12	313	6	333	360	97.90%	16	-	3807780	3808109	330
18_13	316	2	333	358	97.90%	16	-	3807775	3808109	335

Side by Side Alignment

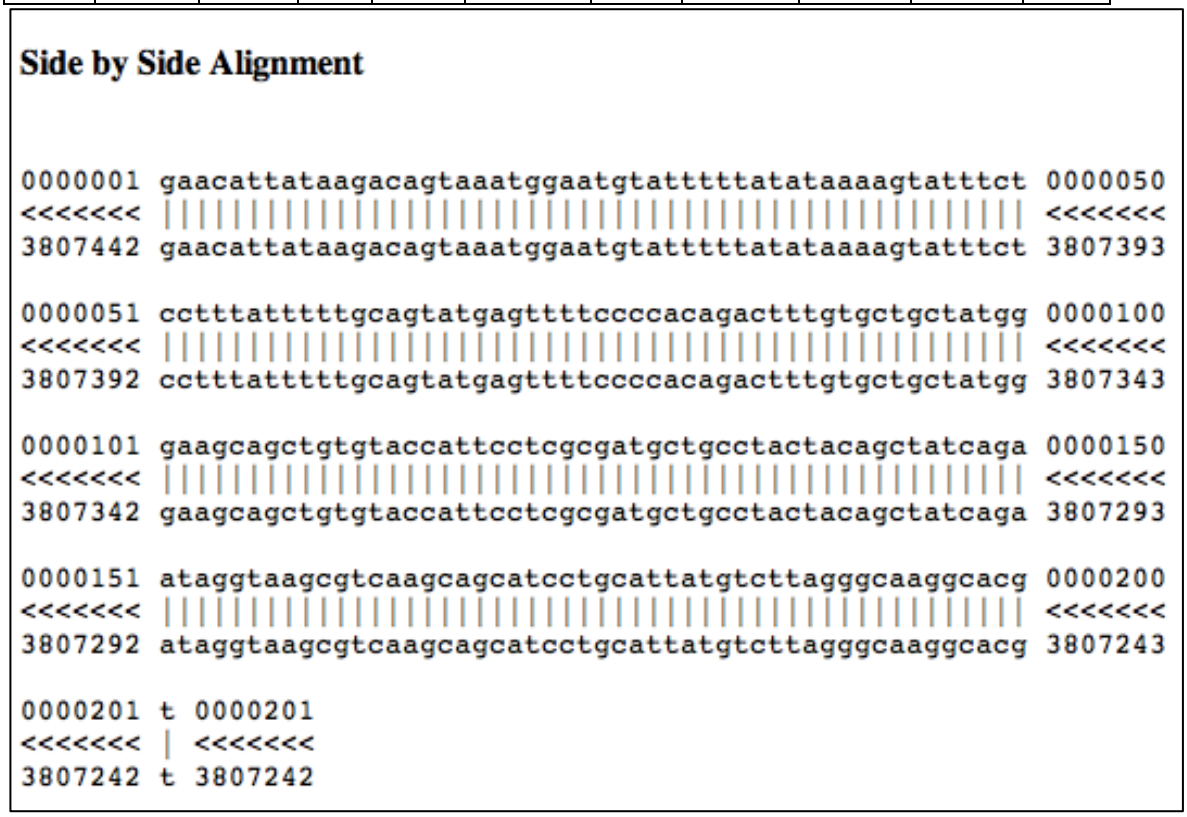
```
0000001 gccagatgagactggcatttggatattgggggattccctatactgagacc 0000050
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3808116 gccagatgagactggcatttggatattgggggattccctatactgagacc 3808067

0000051 attttttttttttaaggactattttgacatcgtaaagaatcccatggac 0000100
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3808066 attttttttttttaaggactattttgacatcgtaaagaatcccatggac 3808017

0000101 ctctccaccatcaagcggaagctggacacagggcaataccaagagccctg 0000150
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3808016 ctctccaccatcaagcggaagctggacacagggcaataccaagagccctg 3807967

0000151 gcagtacgtggacgacgtctggctcatgttcaacaatgcctggctctata 0000200
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807966 gcagtacgtggacgacgtctggctcatgttcaacaatgcctggctctata 3807917

0000201 atcgcaagacatcccgagtctataagttttgcagtaagcttgcagaggtc 0000250
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807916 atcgcaagacatcccgagtctataagttttgcagtaagcttgcagaggtc 3807867

0000251 tttgagcaggaaattgaccctgtcatgcagtcccttggatattgctgtgg 0000300
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807866 tttgagcaggaaattgaccctgtcatgcagtcccttggatattgctgtgg 3807817

0000301 acgcaaggtacagttttaagtttttccggaaagtgaattttcctggttaa 0000350
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807816 acgcaaggtacagttttaagtttttccggaaagtgaattttcctggttaa 3807767

0000351 tccagccagaggggt 0000365
<<<<<<< ||||||||||||||| <<<<<<<
3807766 tccagccagaggggt 3807752
```

Figure A.4.1, showing the alignment result for exon 18 using the expected PCR product for the used primer set.

Table A.4.2 showing the summary Blat results of PCR products for the RTS samples using exon 19 primer set.

	Exon 19 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 19	201	1	201	201	100.00%	16	-	3807242	3807442	201
19_1	183	2	186	205	98.40%	16	-	3807242	3807425	184
19_2	181	3	184	203	100.00%	16	-	3807242	3807424	183
19_3	166	20	185	203	100.00%	16	-	3807277	3807442	166
19_4	172	9	180	199	100.00%	16	-	3807271	3807442	172
19_5										
19_6										
19_7	175	5	179	197	100.00%	16	-	3807268	3807442	175
19_8	144	12	182	199	93.50%	16	-	3807271	3807442	172
19_9	150	15	190	214	89.60%	16	-	3807273	3807439	167
19_10	179	1	179	197	100.00%	16	-	3807264	3807442	179
19_11	172	9	180	199	100.00%	16	-	3807271	3807442	172
19_12	177	2	178	197	100.00%	16	-	3807266	3807442	177

19_13	157	10	188	209	90.50%	16	-	3807273	3807442	170



Side by Side Alignment

```
0000001 gaacattataagacagtaaatggaatgtatttttatataaaagtatttct 0000050
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807442 gaacattataagacagtaaatggaatgtatttttatataaaagtatttct 3807393

0000051 cctttatttttgcagtatgagttttccccacagactttgtgctgctatgg 0000100
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807392 cctttatttttgcagtatgagttttccccacagactttgtgctgctatgg 3807343

0000101 gaagcagctgtgtaccattcctcgcgatgctgcctactacagctatcaga 0000150
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807342 gaagcagctgtgtaccattcctcgcgatgctgcctactacagctatcaga 3807293

0000151 ataggtaagcgtcaagcagcatcctgcattatgtcttagggcaaggcacg 0000200
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3807292 ataggtaagcgtcaagcagcatcctgcattatgtcttagggcaaggcacg 3807243

0000201 t 0000201
<<<<<<< | <<<<<<<
3807242 t 3807242
```

Figure A.4.2, showing the alignment result for exon 19 using the expected PCR product for the used primer set.

Table A.4.3 showing the summary Blat results of PCR products for the RTS samples using exon 20 primer set

	Exon 20 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 20	190	1	190	190	100.00%	16	-	3801677	3801866	190
20_1	183	4	216	349	99.50%	16	-	3801552	3801849	298
20_2	166	6	173	306	100.00%	16	-	3801677	3801846	170
20_3	161	9	169	187	100.00%	16	-	3801706	3801866	161
20_4	161	8	168	185	100.00%	16	-	3801706	3801866	161
20_5	146	11	172	188	98.10%	16	-	3801705	3801866	162
20_6	158	10	167	186	100.00%	16	-	3801709	3801866	158
20_7	154	11	170	188	99.40%	16	-	3801705	3801866	162
20_8	160	8	167	185	100.00%	16	-	3801707	3801866	160
20_9	155	8	170	307	95.00%	16	-	3801707	3801866	160
20_10	151	8	174	193	91.30%	16	-	3801707	3801866	160
20_11	159	10	170	188	99.40%	16	-	3801706	3801866	161
20_12	158	8	167	186	99.40%	16	-	3801707	3801866	160
20_13	146	12	176	192	94.80%	16	-	3801705	3801866	162

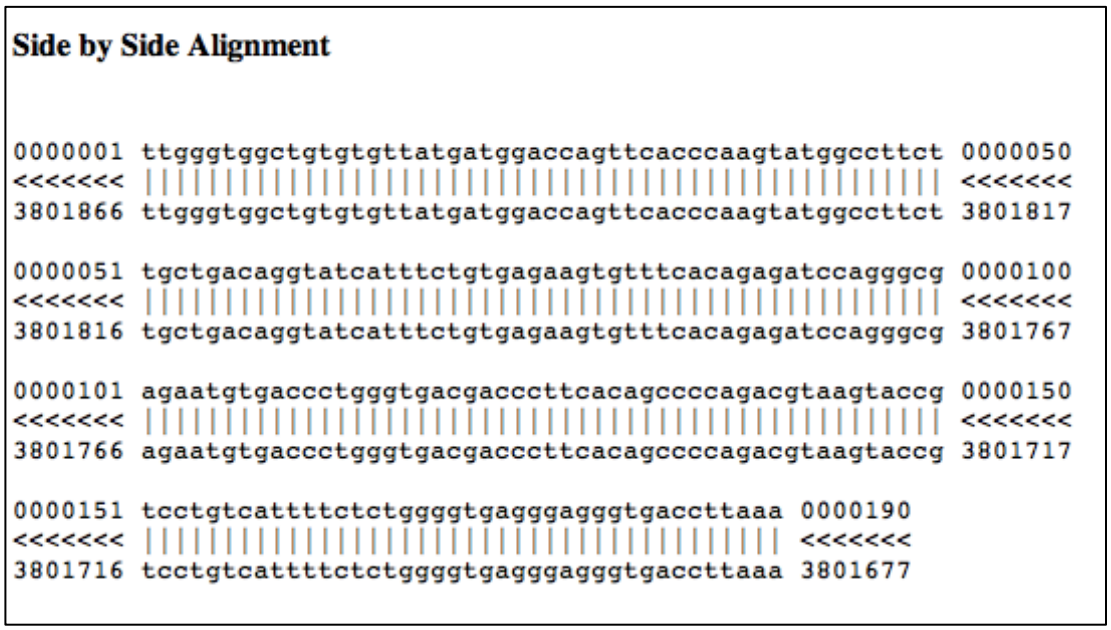


Figure A.4.3, showing the alignment result for exon 20 using the expected PCR product for the used primer set.

Table A.4.4 showing the summary Blat results of PCR products for the RTS samples using exon 21 primer set

	Exon 21 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 21	192	1	192	192	100.00%	16	-	3799578	3799769	192
21_1	166	6	173	192	100.00%	16	-	3799578	3799747	170
21_2	170	6	175	192	100.00%	16	-	3799578	3799747	170
21_3	164	2	166	184	100.00%	16	-	3799604	3799769	166
21_4	157	11	167	278	100.00%	16	-	3799613	3799769	157
21_5										
21_6	164	2	166	282	100.00%	16	-	3799604	3799769	166
21_7	166	2	167	185	100.00%	16	-	3799604	3799769	166
21_8										
21_9	131	10	150	310	97.90%	16	-	3799613	3799758	146
21_10	197	8	291	392	95.40%	16	-	3799611	3808065	8455
21_11	297	1	334	358	97.50%	16	-	3799602	3808111	8510
21_12	164	1	168	186	98.90%	16	-	3799602	3799769	168
21_13	266	4	294	365	98.90%	16	-	3799603	3808065	8463

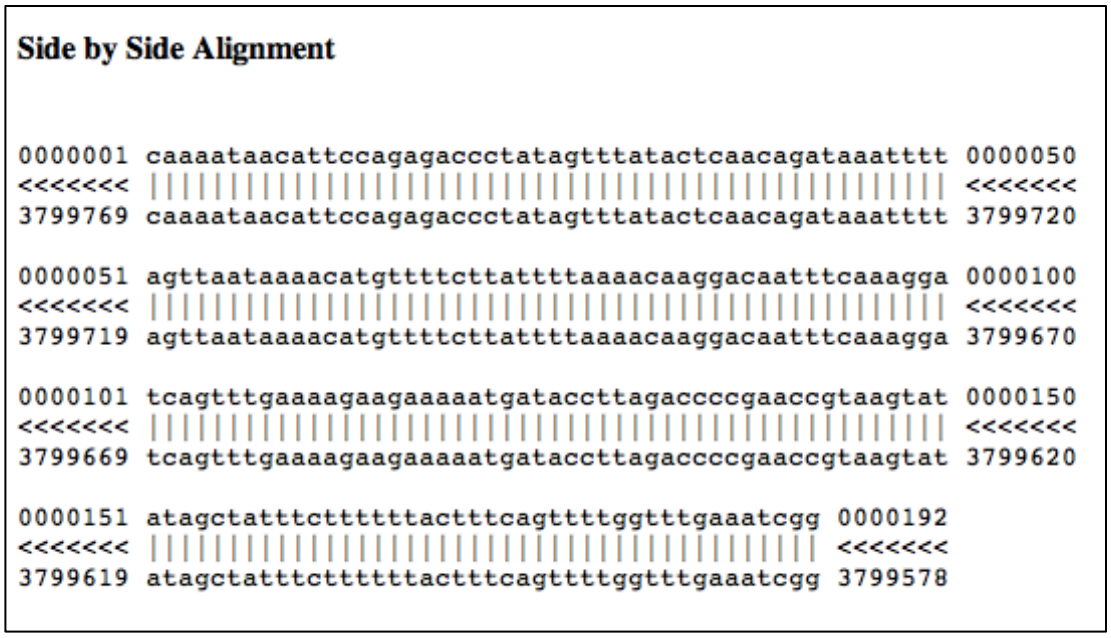


Figure A.4.4, showing the alignment result for exon 21 using the expected PCR product for the used primer set.

Table A.4.5 showing the summary Blat results of PCR products for the RTS samples using exon 22 primer set

	Exon 22 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 22	246	1	246	246	100.00%	16	-	3795191	3795436	246
22_1	229	2	234	386	99.60%	16	-	3795186	3795420	235
22_2	228	5	234	249	99.60%	16	-	3795186	3795415	230
22_3	211	17	227	243	100.00%	16	-	3795228	3795438	211
22_4	219	4	222	237	100.00%	16	-	3795220	3795438	219
22_5	219	4	222	239	100.00%	16	-	3795220	3795438	219
22_6	220	1	224	240	98.30%	16	-	3795216	3795438	223
22_7	218	4	223	240	98.70%	16	-	3795220	3795438	219
22_8										
22_9	203	9	232	491	94.80%	16	-	3795223	3795438	216
22_10	223	1	223	241	100.00%	16	-	3795216	3795438	223
22_11	194	17	230	245	95.20%	16	-	3795228	3795438	211
22_12	207	8	229	290	94.90%	16	-	3795223	3795438	216
22_13	217	1	222	236	99.10%	16	-	3795216	3795438	223

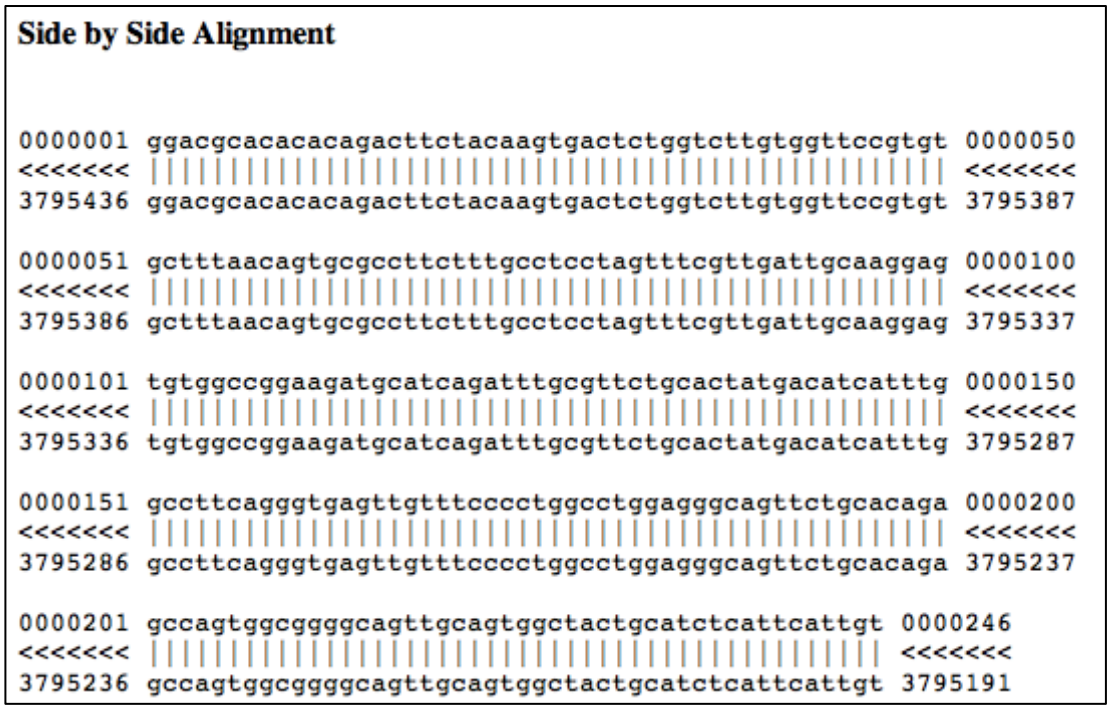


Figure A.4.5, showing the alignment result for exon 22 using the expected PCR product for the used primer set.

Table A.4.6 showing the summary Blat results of PCR products for the RTS samples using exon 23 primer set

	Exon 23 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 23	200	1	200	200	100.00%	16	-	3794830	3795029	200
23_1	173	12	184	203	100.00%	16	-	3794830	3795002	173
23_2	173	13	185	204	100.00%	16	-	3794830	3795002	173
23_3	172	1	174	358	98.30%	16	-	3794857	3795029	173
23_4	170	2	173	381	98.30%	16	-	3794859	3795029	171
23_5	167	2	171	380	99.50%	16	-	3794859	3795029	171
23_6	169	1	173	382	98.90%	16	-	3794857	3795029	173
23_7	169	2	172	368	99.50%	16	-	3794859	3795029	171
23_8	169	2	171	377	100.00%	16	-	3794859	3795029	171
23_9	169	2	172	376	99.50%	16	-	3794859	3795029	171
23_10	166	12	177	194	100.00%	16	-	3794864	3795029	166
23_11	168	5	174	365	99.50%	16	-	3794860	3795029	170
23_12	157	16	180	385	95.80%	16	-	3794867	3795029	163
23_13	167	1	171	368	99.50%	16	-	3794858	3795029	172

Figure A.4.6, showing the alignment result for exon 23 using the expected PCR product for the used primer set.

Table A.4.7 showing the summary Blat results of PCR products for the RTS samples using exon 24 primer set

	Exon 24 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 24	274	1	274	274	100.00%	16	-	3790319	3790592	274
24_1	253	6	264	278	99.30%	16	-	3790306	3790571	266
24_2	251	3	269	283	97.40%	16	-	3790306	3790574	269
24_3	242	11	253	270	100.00%	16	-	3790349	3790592	244
24_4	242	12	254	271	100.00%	16	-	3790349	3790592	244
24_5	206	14	261	277	92.40%	16	-	3790348	3790592	245
24_6	244	10	253	272	100.00%	16	-	3790349	3790592	244
24_7	212	30	258	275	96.60%	16	-	3790363	3790592	230
24_8	230	13	253	272	99.20%	16	-	3790349	3790592	244
24_9	244	1	256	274	97.20%	16	-	3790340	3790592	253
24_10	233	27	259	277	100.00%	16	-	3790360	3790592	233
24_11	232	30	261	280	100.00%	16	-	3790361	3790592	232
24_12	234	26	259	277	100.00%	16	-	3790359	3790592	234
24_13	237	5	253	272	99.60%	16	-	3790342	3790592	251

Figure A.4.7, showing the alignment result for exon 24 using the expected PCR product for the used primer set.

Table A.4.8 showing the summary Blat results of PCR products for the RTS samples using exon 25 primer set

	Exon 25 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 25	258	1	258	258	100.00%	16	-	3789534	3789791	258
25_1	231	2	235	253	99.60%	16	-	3789534	3789769	236
25_2	226	10	236	257	100.00%	16	-	3789536	3789764	229
25_3										
25_4										
25_5										
25_6										
25_7										
25_8										
25_9										
25_10										
25_11										
25_12										
25_13										

Figure A.4.8, showing the alignment result for exon 25 using the expected PCR product for the used primer set.

Table A.4.9 showing the summary Blat results of PCR products for the RTS samples using exon 26 primer set

	Exon 26 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 26	308	1	308	308	100.00%	16	-	3788447	3788754	308
26_1	284	10	294	314	100.00%	16	-	3788447	3788732	286
26_2	282	11	294	314	100.00%	16	-	3788447	3788732	286
26_3	282	1	282	300	100.00%	16	-	3788473	3788754	282
26_4	282	1	282	301	100.00%	16	-	3788473	3788754	282
26_5	280	1	281	353	100.00%	16	-	3788443	3788754	312
26_6	278	4	281	313	100.00%	16	-	3788477	3788754	278
26_7	276	1	281	303	99.30%	16	-	3788473	3788754	282
26_8	282	1	282	302	100.00%	16	-	3788473	3788754	282
26_9	278	4	281	312	100.00%	16	-	3788477	3788754	278
26_10	288	4	319	354	99.70%	16	-	3788477	3788926	450
26_11	264	8	286	307	98.20%	16	-	3788476	3788754	279
26_12	278	4	281	299	100.00%	16	-	3788477	3788754	278
26_13	282	1	282	300	100.00%	16	-	3788473	3788754	282

Side by Side Alignment

```
0000001 ttccagggtgttgtttgttgcttgtgtttgatttcagatttgagggatac 0000050
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3788754 ttccagggtgttgtttgttgcttgtgtttgatttcagatttgagggatac 3788705

0000051 cctgagttaaacatgtgcctccttcccacaggcgtgtgtacatttcttat 0000100
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3788704 cctgagttaaacatgtgcctccttcccacaggcgtgtgtacatttcttat 3788655

0000101 ctggatagtattcatttcttccggccacgttgcctccgcacagccgttta 0000150
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3788654 ctggatagtattcatttcttccggccacgttgcctccgcacagccgttta 3788605

0000151 ccatgagatccttattggatatttagagtatgtgaagaaattagggtgag 0000200
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3788604 ccatgagatccttattggatatttagagtatgtgaagaaattagggtgag 3788555

0000201 tttggattaaattatttggaactagtaagaacccttttatttttactgta 0000250
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3788554 tttggattaaattatttggaactagtaagaacccttttatttttactgta 3788505

0000251 tgtttattccgtgaaataatgggtattttaagttttatgcgttttttattt 0000300
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3788504 tgtttattccgtgaaataatgggtattttaagttttatgcgttttttattt 3788455

0000301 ttccatcc 0000308
<<<<<<< |||||||| <<<<<<<
3788454 ttccatcc 3788447
```

Figure A.4.9, showing the alignment result for exon 26 using the expected PCR product for the used primer set.

Table A.4.10 showing the summary Blat results of PCR products for the RTS samples using exon 27 primer set

	Exon 27 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 27	301	1	301	301	100.00%	16	-	3786577	3786877	301
27_1	277	7	284	304	100.00%	16	-	3786577	3786855	279
27_2	281	2	284	304	100.00%	16	-	3786577	3786863	287
27_3	274	5	278	389	100.00%	16	-	3786604	3786877	274
27_4	262	22	301	306	96.60%	16	-	3786603	3786876	274
27_5	227	7	278	295	94.90%	16	-	3786604	3786875	272
27_6	272	1	277	620	99.30%	16	-	3786600	3786877	278
27_7	260	1	278	498	96.80%	16	-	3786600	3786877	278
27_8	274	1	277	463	99.70%	16	-	3786600	3786877	278
27_9	261	24	288	306	98.10%	16	-	3786615	3786877	263
27_10	231	17	287	304	96.30%	16	-	3786610	3786877	268
27_11	177	5	257	279	99.00%	16	-	3786604	3786876	273
27_12	274	1	278	508	99.30%	16	-	3786600	3786877	278
27_13	272	1	277	620	99.30%	16	-	3786600	3786877	278

Side by Side Alignment

```
0000001 cttaaaggcagggccgatttcactggcacgttcatctgacgtgtgtgcgt 0000050
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786877 cttaaaggcagggccgatttcactggcacgttcatctgacgtgtgtgcgt 3786828

0000051 gggtcctgcaggtatgtgacagggcacatctgggcctgtcctccaagtga 0000100
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786827 gggtcctgcaggtatgtgacagggcacatctgggcctgtcctccaagtga 3786778

0000101 aggagatgattacatcttccattgccacccacctgatcaaaaaataccca 0000150
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786777 aggagatgattacatcttccattgccacccacctgatcaaaaaataccca 3786728

0000151 agccaaaacgactgcaggagtggtacaaaaagatgctggacaaggcgttt 0000200
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786727 agccaaaacgactgcaggagtggtacaaaaagatgctggacaaggcgttt 3786678

0000201 gcagagcggatcatccatgactacaaggtacccggcgctgaaggaaggtg 0000250
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786677 gcagagcggatcatccatgactacaaggtacccggcgctgaaggaaggtg 3786628

0000251 gtggcttttgtcacaactgaggggaggatatttgtgtgccttttttcttgc 0000300
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786627 gtggcttttgtcacaactgaggggaggatatttgtgtgccttttttcttgc 3786578

0000301 a 0000301
<<<<<<< | <<<<<<<
3786577 a 3786577
```

Figure A.4.10, showing the alignment result for exon 27 using the expected PCR product for the used primer set.

Table A.4.11 showing the summary Blat results of PCR products for the RTS samples using exon 28 primer set

	Exon 28 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 28	327	1	327	327	100.00%	16	-	3786000	3786326	327
28_1	303	9	313	539	99.10%	16	-	3786000	3786303	304
28_2	304	9	312	332	100.00%	16	-	3786000	3786303	304
28_3	300	8	311	333	98.40%	16	-	3786025	3786326	302
28_4	297	8	304	322	100.00%	16	-	3786030	3786326	297
28_5	299	10	308	401	100.00%	16	-	3786028	3786326	299
28_6	304	2	307	437	99.70%	16	-	3786021	3786326	306
28_7	302	1	305	476	99.70%	16	-	3786021	3786326	306
28_8	299	7	305	472	100.00%	16	-	3786028	3786326	299
28_9										
28_10	301	7	307	400	100.00%	16	-	3786026	3786326	301
28_11	299	6	304	474	100.00%	16	-	3786028	3786326	299
28_12	300	7	306	435	100.00%	16	-	3786027	3786326	300
28_13	297	9	305	476	100.00%	16	-	3786030	3786326	297

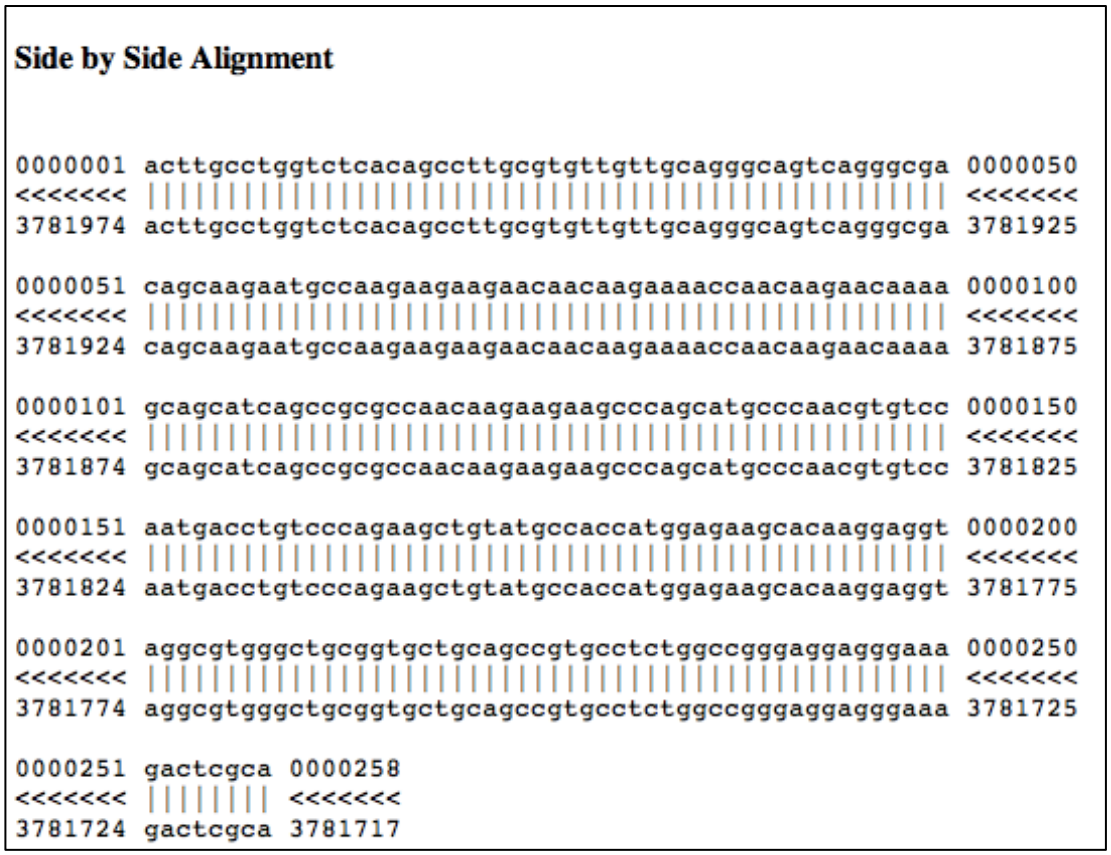
Side by Side Alignment

```
0000001 cacacatgcatgggactctgccacaccatggtgaggggcacgcttgcgtg 0000050
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786326 cacacatgcatgggactctgccacaccatggtgaggggcacgcttgcgtg 3786277

0000051 ggggtctgtcgcaccatggtgggggggcacgcgtgcatggccctcatctca 0000100
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786276 ggggtctgtcgcaccatggtgggggggcacgcgtgcatggccctcatctca 3786227

0000101 ctgttgtgctttgcccactcaggatattttcaaacaagcaactgaagaca 0000150
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786226 ctgttgtgctttgcccactcaggatattttcaaacaagcaactgaagaca 3786177

0000151 ggctcaccagtgccaaggaactgccctattttgaaggtgatttctggccc 0000200
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786176 ggctcaccagtgccaaggaactgccctattttgaaggtgatttctggccc 3786127

0000201 aatgtgttagaagagagcattaaggaactagaacaagaagaagaggagag 0000250
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786126 aatgtgttagaagagagcattaaggaactagaacaagaagaagaggagag 3786077

0000251 gaaaaaggaagagagcactgcagccagtgaaaccactgaggtacagaccc 0000300
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3786076 gaaaaaggaagagagcactgcagccagtgaaaccactgaggtacagaccc 3786027

0000301 ttctctgagctccattgcccacgtgtc 0000327
<<<<<<< ||||||||||||||||||||||||||| <<<<<<<
3786026 ttctctgagctccattgcccacgtgtc 3786000
```

Figure A.4.11, showing the alignment result for exon 28 using the expected PCR product for the used primer set.

Table A.4.12 showing the summary Blat results of PCR products for the RTS samples using exon 29 primer set

	Exon 29 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 29	258	1	258	258	100.00%	16	-	3781717	3781974	258
29_1	234	4	238	482	100.00%	16	-	3781716	3781951	236
29_2	232	6	241	473	99.20%	16	-	3781716	3781951	236
29_3	234	2	235	484	100.00%	16	-	3781741	3781974	234
29_4	245	7	445	494	94.00%	16	-	3781744	3782103	360
29_5	234	2	235	482	100.00%	16	-	3781741	3781974	234
29_6	234	1	236	481	99.60%	16	-	3781739	3781974	236
29_7	233	5	237	470	100.00%	16	-	3781742	3781974	233
29_8	235	1	238	461	98.40%	16	-	3781739	3781974	236
29_9	104	7	205	329	84.20%	16	-	3786025	3786220	196
29_10	231	6	236	469	100.00%	16	-	3781744	3781974	231
29_11	229	8	236	471	100.00%	16	-	3781746	3781974	229
29_12	234	1	236	466	99.60%	16	-	3781739	3781974	236
29_13	233	5	237	479	100.00%	16	-	3781742	3781974	233



Side by Side Alignment

```
0000001 acttgcctggtctcacagccttgcgtgttgttgcagggcagtcagggcga 0000050
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781974 acttgcctggtctcacagccttgcgtgttgttgcagggcagtcagggcga 3781925

0000051 cagcaagaatgccaagaagaagaacaacaagaaaaccaacaagaacaaaa 0000100
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781924 cagcaagaatgccaagaagaagaacaacaagaaaaccaacaagaacaaaa 3781875

0000101 gcagcatcagccgcgccaacaagaagaagcccagcatgcccaacgtgtcc 0000150
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781874 gcagcatcagccgcgccaacaagaagaagcccagcatgcccaacgtgtcc 3781825

0000151 aatgacctgtcccagaagctgtatgccaccatggagaagcacaaggaggt 0000200
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781824 aatgacctgtcccagaagctgtatgccaccatggagaagcacaaggaggt 3781775

0000201 aggcgtgggctgcggtgctgcagccgtgcctctggccgggaggagggaaa 0000250
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781774 aggcgtgggctgcggtgctgcagccgtgcctctggccgggaggagggaaa 3781725

0000251 gactcgca 0000258
<<<<<<< |||||||| <<<<<<<
3781724 gactcgca 3781717
```

Figure A.4.12, showing the alignment result for exon 29 using the expected PCR product for the used primer set.

Table A.4.13 showing the summary Blat results of PCR products for the RTS samples using exon 30 primer set

	Exon 30 results									
	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
exon 30	397	1	397	397	100.00%	16	-	3781133	3781529	397
30_3	368	7	374	612	100.00%	16	-	3781162	3781529	368
30_4	81	5	85	395	100.00%	16	-	3781160	3781240	81
30_5										
30_6	373	2	374	583	100.00%	16	-	3781157	3781529	373
30_7	370	2	375	393	99.00%	16	-	3781157	3781529	373
30_8										
30_9	374	1	374	614	100.00%	16	-	3781156	3781529	374
30_10	372	3	374	546	100.00%	16	-	3781158	3781529	372
30_11	373	1	373	548	100.00%	16	-	3781157	3781529	373
30_12	369	7	377	395	99.20%	16	-	3781160	3781529	370
30_13	374	1	374	612	100.00%	16	-	3781156	3781529	374

Side by Side Alignment

```
0000001 accactggaggtgccatgtcccttgtgtgggactaaagcccctcctctcc 0000050
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781529 accactggaggtgccatgtcccttgtgtgggactaaagcccctcctctcc 3781480

0000051 tgcaggtcttcttcgtgatccacctgcacgctgggcctgtcatcaacacc 0000100
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781479 tgcaggtcttcttcgtgatccacctgcacgctgggcctgtcatcaacacc 3781430

0000101 ctgccccccatcgtcgaccccgaccccctgctcagctgtgacctcatgga 0000150
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781429 ctgccccccatcgtcgaccccgaccccctgctcagctgtgacctcatgga 3781380

0000151 tgggcgcgacgccttcctcaccctcgccagagacaagcactgggagttct 0000200
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781379 tgggcgcgacgccttcctcaccctcgccagagacaagcactgggagttct 3781330

0000201 cctccttgcgccgctccaagtggtccacgctctgcatgctggtggagctg 0000250
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781329 cctccttgcgccgctccaagtggtccacgctctgcatgctggtggagctg 3781280

0000251 cacacccagggccaggaccgctttgtctacacctgcaacgagtgcaagca 0000300
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781279 cacacccagggccaggaccgctttgtctacacctgcaacgagtgcaagca 3781230

0000301 ccacgtggagacgcgctggcactgcactgtgtgcgaggtaggccctgccc 0000350
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781229 ccacgtggagacgcgctggcactgcactgtgtgcgaggtaggccctgccc 3781180

0000351 ccacccccaccccccacagccggcctggggtctgacgaagcatcctgt 0000397
<<<<<<< |||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<
3781179 ccacccccaccccccacagccggcctggggtctgacgaagcatcctgt 3781133
```

Figure A.4.13, showing the alignment result for exon 29 using the expected PCR product for the used primer set.

Appendix for Chapter 5

Table 1. SALSA MLPA P036-E1 Human Telomere-3 probemix

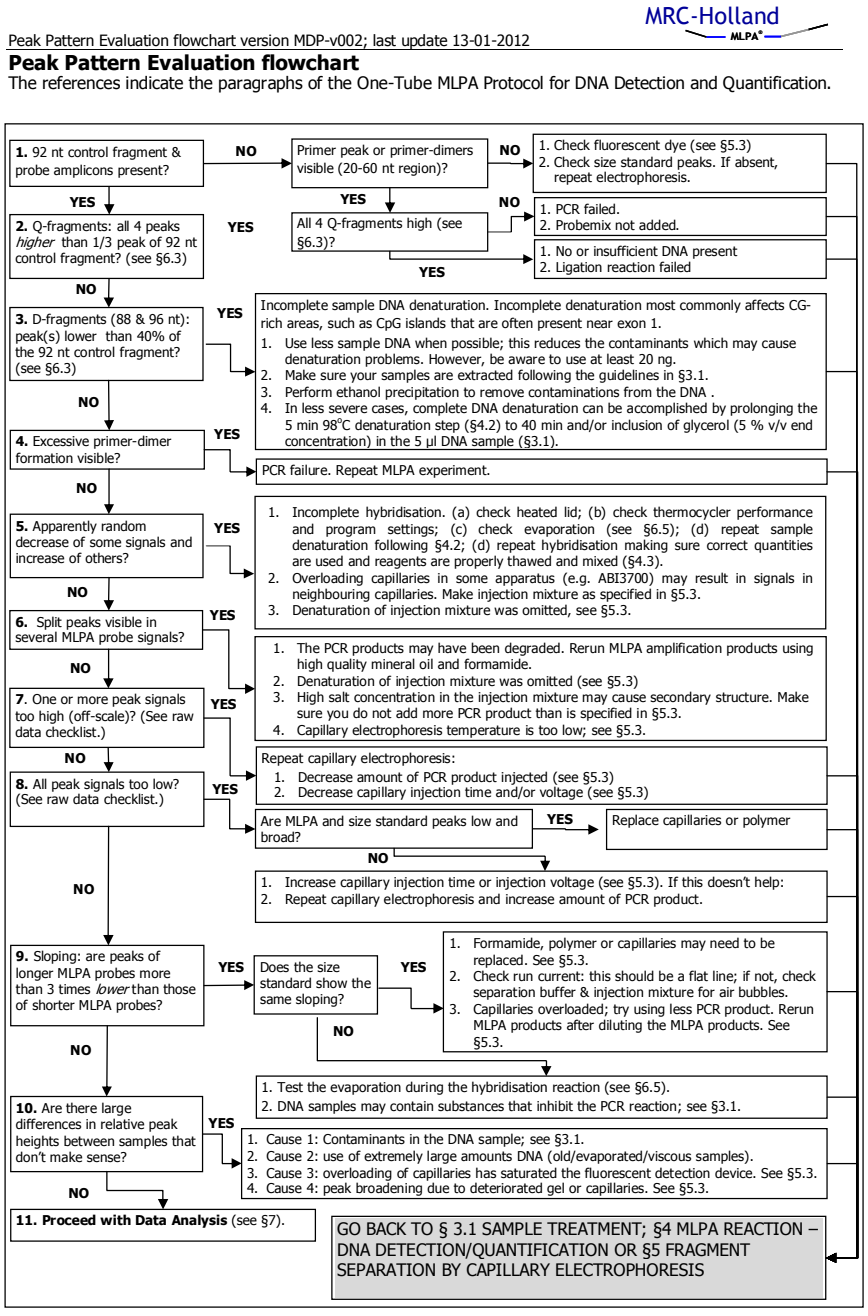
Length (nt)	Chromosomal position	Gene detected	SALSA MLPA probe	MapView build 36 position
64-70-76-82	Q-fragments: DNA quantity; only visible with less than 100 ng sample DNA			
88-92-96	D-fragments: Low signal of 88 or 96 nt fragment indicates incomplete denaturation			
100	X-fragment: Specific for the X chromosome			
105	Y-fragment: Specific for the Y chromosome			
118	Y-fragment: Specific for the Y chromosome			
130 §	1p	TNFRSF4	02269-L01761	01-001.14
136	2p	ACP1	02274-L08758	02-000.25
142	3p	CHL1	01721-L01329	03-000.34
151	4p	PIGG (FLJ20265)	02005-L02047	04-000.50
158	5p	PDCD6	01723-L01327	05-000.37
166	6p	IRF4	01724-L02048	06-000.34
172	7p	ADAP1 (CENTA1)	02275-L02049	07-000.93
179	8p	FBXO25	02397-L01845	08-000.40
186	9p	DMRT1	01727-L02050	09-000.84
193	10p	DIP2C (KIAA0934)	02277-L01768	10-000.48
202	11p	RIC8A (RIC-8)	03315-L02733	11-000.20
208	12p	SLC6A12	02276-L01767	12-000.17
219 +	13q-cen	PSPC1	02399-L01847	13-019.24 (Acrocentric chromosome)
227 +	14q-cen	CCNB1IP1 (HEI10)	01732-L01318	14-019.86 (Acrocentric chromosome)
235 +	15q-cen	MKRN3	07291-L08858	15-021.36 (Acrocentric chromosome)
242	16p	POLR3K	01734-L01316	16-000.04
250	17p	RPH3AL	01735-L01315	17-000.17
258	18p	USP14	01736-L02051	18-000.19
265	19p	CDC34	01737-L01313	19-000.49
274	20p	SOX12	02396-L01844	20-000.26
283 +	21q-cen	RBM11	01739-L01311	21-014.51 (Acrocentric chromosome)
289 +	22q-cen	BID	01740-L01310	22-016.61 (Acrocentric chromosome)
298	Xp/Yp (PAR1)	SHOX	01148-L01331	X/Y-000.52 (PAR1 region)
307	1q	SH3BP5L (KIAA1720)	02392-L02149	01-247.08 (0.2 Mb from telomere)
313	2q	CAPN10	01742-L01308	02-241.18 (1.6 Mb from telomere)
322	3q	BDH1	02013-L02052	03-198.76 (0.7 Mb from telomere)
330 §	4q	TRIML2	12050-L11446	04-189.26 (2.0 Mb from telomere)
337	5q	GNB2L1	03319-L02737	05-180.60 (0.2 Mb from telomere)
346	6q	PSMB1	01746-L01304	06-170.69 (0.5 Mb from telomere)
355	7q	VIPR2	01747-L01303	07-158.60 (0.3 Mb from telomere)
361	8q	ZC3H3 (KIAA0150)	01748-L01302	08-144.69 (1.6 Mb from telomere)
372	9q	EHMT1	08205-L08170	09-139.83 (0.2 Mb from telomere)
379	10q	PAOX (PAO)	09142-L09953	10-135.05 (0.2 Mb from telomere)
386	11q	NCAPD3 (KIAA0056)	01751-L01299	11-133.60 (1.2 Mb from telomere)
395	12q	ZNF10	02687-L02154	12-132.24 (0.2 Mb from telomere)
402	13q	F7	01753-L01297	13-112.82 (1.3 Mb from telomere)
411	14q	MTA1	02778-L02201	14-105.00 (1.3 Mb from telomere)
418	15q	ALDH1A3	01755-L01295	15-099.26 (1.0 Mb from telomere)
426	16q	GAS8 (GAS11)	03201-L02669	16-088.63 (0.2 Mb from telomere)
434	17q	TBCD	01757-L01293	17-078.45 (0.5 Mb from telomere)
441	18q	C18orf22 (FLJ21172)	01758-L01292	18-075.90 (0.2 Mb from telomere)
450	19q	CHMP2A (BC-2)	09143-L10626	19-063.75 (0.9 Mb from telomere)
458	20q	OPRL1	02688-L02884	20-062.19 (0.2 Mb from telomere)
466	21q	PRMT2 (HMT1)	02586-L02059	21-046.89 (0.1 Mb from telomere)
475	22q	RABL2B	01762-L08761	22-049.55 (0.1 Mb from telomere)
483	Xq/Yq (PAR2)	VAMP7 (SYBL1)	01763-L02150	X/Y-154.78 (PAR2; 0.1 Mb from telom.)

Table 1. SALSA MLPA P070-B2 Human Telomere-5 probemix

Length (nt)	Chromosomal position	Gene detected	SALSA MLPA probe	MapView build 36 position
64-70-76-82	Q-fragments: DNA quantity; only visible with less than 100 ng sample DNA			
88-92-96	D-fragments: Low signal of 88 or 96 nt fragment indicates incomplete denaturation (autosomal)			
100	X-fragment: Specific for the X chromosome (AMOT gene ; X-111.95)			
105	Y-fragment: Specific for the Y chromosome (UTY gene ; Y-013.98)			
118	Y-fragment: Specific for the Y chromosome (DDX3Y gene ; Y-013.54)			
132	1q	SH3BP5L (KIAA1720)	04084-L03605	01-247.08
139	2q	ATG4B (=APG4B)	02781-L03168	02-242.25
145	3q	KIAA0226	02690-L02842	03-198.88
152	4q	FRG1	02691-L02843	04-191.10
160	5q	GNB2L1	02790-L02232	05-180.60
166	6q	TBP	02694-L02844	06-170.71
172	7q	VIPR2	02793-L03167	07-158.63
179	8q	RECQL4	02695-L00610	08-145.71
186	9q	EHMT1	02792-L02846	09-139.78
193	10q	ECHS1	02696-L02847	10-135.03
202	11q	IGSF9B (=KIAA1030)	02697-L02848	11-133.29
211	12q	ZNF10	02686-L02849	12-132.24
218	13q	CDC16	02698-L00753	13-114.03
226	14q	MTA1	02699-L02850	14-104.99
233	15q	TM2D3 (=FLJ22604)	02701-L02851	15-100.01
241	16q	GAS8 (=GAS11)	02702-L00734	16-088.64
250	17q	SECTM1	02703-L03169	17-077.87
258	18q	CTDP1	02704-L03607	18-075.58
265	19q	CHMP2A (=BC-2)	02705-L02853	19-063.76
274	20q	UCKL1 (=FLJ20517)	02706-L00642	20-062.05
281	21q	S100B	02587-L02854	21-046.85
290	22q	ARSA	02707-L00661	22-049.41
298	X/Yq (PAR2)	VAMP7 (=SYBL1)	02708-L02855	X-154.82 + Y-057.68 (PAR region)
306	1p	TNFRSF18	02270-L01762	01-001.13
315	2p	ACP1	02709-L02856	02-000.27
323	3p	CHL1	02896-L02363	03-000.34
329	4p	PIGG	14440-L16146	04-000.51
337	5p	CCDC127 (=LOC133957)	02791-L02233	05-000.26
346	6p	IRF4	04077-L03462	06-000.34
355	7p	UNC84A	02780-L02857	07-000.84
362	8p	FBXO25	02715-L00973	08-000.40
370	9p	DOCK8 (=FLJ00026)	02716-L00688	09-000.38
379	10p	ZMYND11 (=BS69)	05180-L16343	10-000.22
387	11p	BET1L	02784-L02226	11-000.20
393	12p	JARID1A (=RBBP2)	02787-L02229	12-000.29
402 +	"13p"	PSPC1	02717-L03608	13-019.25 (Acrocentric)
409 +	"14p"	PARP2 (=ADPRTL2)	02718-L00732	14-019.90 (Acrocentric)
418 +	"15p"	NDN	04026-L01542	15-021.48 (Acrocentric)
427	16p	DECR2	02720-L00648	16-000.40
436	17p	RPH3AL	04081-L03465	17-000.18
444	18p	THOC1	02789-L02231	18-000.20
450	19p	PPAP2C	03501-L02880	19-000.23
459	20p	ZCCHC3 (=FLJ22115)	02723-L00641	20-000.23
466 +	"21p"	HSPA13 (=STCH)	02724-L00334	21-014.68 (Acrocentric)
478 +	"22p"	IL17RA	02725-L16344	22-015.96 (Acrocentric)
484	X/Yp (PAR1)	SHOX	03714-L16345	X/Y-000.52 (PAR region)

+ The 13, 14, 15, 21 & 22 "p" probes in fact target the q arm, as these chromosomes are acrocentric.
The lengths of some probes in this description have been adjusted as compared to previous versions.

MLPA Peak Pattern evaluation flow chart



Appendix for chapter 6

Cytogenetic array results of selection method 2

Family	File	CN State	Type	Chromosome	Min	Max	Size (kbp)	Marker Count	Confidence	Genes in the region (number of genes followed by gene name)	DGV	OMIM ®
1	100824-01	3	Gain	8	142450527	144144519	1694	184	0.88	**2 genes**	138	
	10120101	4	Gain	8	142478604	142663923	185.32	7	0.93	FLJ43860	42	
	10120101	4	Gain	8	143308105	143373178	65.073	11	0.91	TSNARE1		
1	100824-01	3	Gain	8	144261481	145817594	1556.1	147	0.88	**42 genes**	170	
										ADCK5, BOP1, C8orf30A, C8ORFK29, CPSF1, CYC1, CYHR1, DGAT1, EPPK1, EXOSC4, FBXL6, FOXH1, GPAA1, GPR172A, GPT, GRINA, HEATR7A, HSF1, KIAA1875, KIFC2, LRRC14, LRRC24, MAF1, MFSD3, MIR1234, MIR661, MIR939, NFKBIL2, OPLAH, PARP10, PLEC, PLEC, PPP1R16A, RECQL4, SCRT1, SCXA, SCXB, SCXB, SHARPIN, SLC39A4, SPATC1,		
	10120101	3	Gain	8	144942997	145751248	808.25	65	0.88	VPS28,	78	
2	100824-02	3	Gain	8	143272874	143737908	465.03	53	0.88	**2 genes**		
	10120602	4	Gain	8	143180313	143373178	192.87	14	0.96	NCRNA00051, TSNARE1		
2	10120602	3	Gain	8	144874789	145789811	915.02	73	0.89	**19 genes**	50	
										ADCK5, BOP1, C8ORFK29, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, GPR172A, HSF1, KIFC2, MIR1234, MIR939, NFKBIL2, SCRT1, SCXA, SCXB, SLC39A4,		
	100824-02	3	Gain	8	145484442	145710235	225.79	16	0.92	VPS28,	17	
3	B0921	4	Gain	8	143144004	143391651	247.65	19	0.94	**2 genes**	6	
	10120102	3	Gain	8	142450527	144135307	1684.8	179	0.88	NCRNA00051, TSNARE1	71	
3	B0921	3	Gain	8	144874789	145817594	942.81	79	0.87	**46 genes**	50	
	10120102	3	Gain	8	144221252	145833715	1612.5	151	0.88	ADCK5, ARHGAP39, BOP1, C8orf30A, C8ORFK29, CPSF1, CYC1, CYHR1, DGAT1, EPPK1, EXOSC4, FBXL6, FOXH1, GPAA1, GPR172A, GPT, GRINA, HEATR7A, HSF1, KIAA1875, KIFC2, LRRC14, LRRC24, MAF1, MFSD3, MGC70857, MIR1234, MIR661, MIR937, MIR939, NFKBIL2, NRBP2, OPLAH, PARP10, PLEC, PPP1R16A, PUF60, RECQL4, SCRIB, SCRT1, SCXA, SCXB, SHARPIN, SLC39A4, SLC39A4, SPATC1, VPS28	173	603780 RAPADILINO_syndrome(266280):NM_004260 (RECQL4), 607059 Acrodermatitis_enteropathica(201100):NM_130849 (SLC39A4), 603780 Baller-Gerold_syndrome(218600):NM_004260 (RECQL4), 607059 Acrodermatitis_enteropathica(201100):NM_017767 (SLC39A4), 603780 Rothmund-Thomson_syndrome(268400):NM_004260 (RECQL4)
9	10121501	1	Loss	2	64673003	65006890	333.89	150	0.864	**3 genes**	3	
	10121502	1	Loss	2	64656408	64894680	238.27	148	0.877	AFTPH, HSPC159, SERTAD2	3	
9	10121501	1	Loss	2	162086299	162297588	211.29	124	0.879	**2 genes**	5	
	10121502	1	Loss	2	162173421	162390737	217.32	160	0.892	TBR1, PSMD14	6	
9	10121501	1	Loss	3	142669503	142898579	229.08	121	0.907	**3 genes**	4	
	10121502	1	Loss	3	142650099	142883928	233.83	114	0.88	CHST2, SR140, PAQR9	2	
9	10121501	3	Gain	10	123262821	123655949	393.13	489	0.924	**2 genes**	4	90 OMIM entries for (FGFR2) for the following conditions Apert_syndrome(101200), Craniofacial-skeletal-dermatologic_dysplasia___(0), Crouzon_syndrome(123500), Gastric_cancer_somatic(137215), Jackson-Weiss_syndrome(123150), Beare-Stevenson_cutis_gyrata_syndrome(123790), Craniosynostosis(0),Antley-Bixler_syndrome(207410), Pfeiffer_syndrome(101600), Saethre-Chotzen_syndrome(101400)
	10121502	3	Gain	10	123262821	123665014	402.19	500	0.935	ATE1, FGFR2	4	
9	10121501	2	Gain	Y	8242192	8637819	395.63	126	0.882	**2 genes**		
	10121502	2	Gain	Y	8469846	8754439	284.59	65	0.852	TTTY18, TTTY19		
11	10191003	3	Gain	8	144869301	145751248	881.95	71	0.886	**44 genes**	45	
	10120601	3	Gain	8	144869301	145751248	881.95	71	0.891	ADCK5, BOP1, C8orf30A, C8ORFK29, CPSF1, CYC1, CYHR1, DGAT1, EPPK1, EXOSC4, FBXL6, FOXH1, GPAA1, GPR172A, GPT, GRINA, HEATR7A, HSF1, KIAA1875, KIFC2, LRRC14, LRRC24, MAF1, MFSD3, MIR1234, MIR661, MIR937, MIR939, NFKBIL2, NRBP2, OPLAH, PARP10, PLEC, PPP1R16A, PUF60, RECQL4, SCRIB, SCRT1, SCXA, SCXB, SHARPIN, SLC39A4, SPATC1, VPS28	45	607059 Acrodermatitis_enteropathica(201100):NM_017767 (SLC39A4), 603780 Rothmund-Thomson_syndrome(268400):NM_004260 (RECQL4), 603780 Baller-Gerold_syndrome(218600):NM_004260 (RECQL4), 603780 RAPADILINO_syndrome(266280):NM_004260 (RECQL4)
11	10191003	4	Gain	8	143308105	143373178	65.073	11	0.956	**1 gene**		
	10120601	3	Gain	8	142130557	144133739	2003.2	216	0.884	TSNARE1	80	610613 Aldosteronism_glucocorticoid-remediable(103900):NM_000497 (CYP11B1), 610613 Adrenal_hyperplasia_congenital_due_to_11-beta-hydroxylase_deficiency(202010):NM_001026213 (CYP11B1), 606119 Meleda_disease(248300):NM_020427 (SLURP1), 610613 Aldosteronism_glucocorticoid-remediable(103900):NM_001026213 (CYP11B1), 124080 Hypoaldosteronism_congenital_due_to_CMO_I_deficiency(203400):NM_000498 (CYP11B2), 124080 Aldosterone_to_renin_ratio_raised___(0):NM_000498 (CYP11B2), 124080 Low_renin_hypertension(0):NM_000498 (CYP11B2)
11	10191003	3	Gain	15	20175623	22752399	2576.8	54	0.885	**13 genes**	269	
										BCL8, CXADRP2, GOLGA6L1, GOLGA6L6, GOLGA8C, GOLGA8DP, LOC646214, LOC727924, NF1P1, OR4M2,		
	10120601	3	Gain	15	20175623	22750488	2574.9	53	0.921	OR4N3P, OR4N4, POTEB	269	

Cytogenetic array results of selection method 3

Family number	Sample	CN State	Type	Chromosome	Min	Max	Size (kbp)	Mean Marker Distance	Gene list matching Cooper list	reported CNV type	OMIM ®
2	10120602	3	Gain	1	3213490	3996253	782.763	7906	**14 genes:** ARHGEF16, C1orf174, CCDC27, DFFB, KIAA0495, KIAA0562, LOC100133612, LOC388588, LRRC47, MEGF6, MIR551A, TP73, TPRG1L, WDR8	Loss	
3	10120102	3	Gain	1	2301690	3807388	1505.7	8703	**25 genes**: ACTRT2, ARHGEF16, C1orf174, C1orf93, CCDC27, DFFB, FLJ42875, HES5, KIAA0495, KIAA0562, LOC388588, LRRC47, MEGF6, MIR551A, MMEL1, MORN1, PANK4, PEX10, PLCH2, PRDM16, RER1, TNFRSF14, TP73, TPRG1L, WDR8	Loss	
3	10120102	3	Gain	9	135998735	140489681	4490.95	55443	**84 genes:** ABCA2, AGPAT2, ANAPC2, C8G, C9orf116, C9orf139, C9orf140, C9orf142, C9orf163, C9orf167, C9orf173, C9orf69, C9orf86, CAMSAP1, CARD9, CLIC3, COBRA1, DNLZ, DPP7, EDF1, EGFL7, ENTPD2, ENTPD8, EXD3, FAM166A, FAM69B, FBXW5, FCN1, FCN2, FUT7, GLT6D1, GPSM1, GRIN1, INPP5E, KIAA0649, KIAA1984, LCN10, LCN12, LCN15, LCN6, LCN8, LCN9, LCNL1, LHX3, LOC100131193, LOC100289341, LOC26102, LRRC26, MAMDC4, MAN1B1, MIR126, MRPL41, MRPS2, NACC2, NDOR1, NELF, NOTCH1, NOXA1, NPDC1, NRARP, OLFM1, PAEP, PHPT1, PMPCA, PNPLA7, QSOX2, RNF208, RNU6ATAC, SDCCAG3, SEC16A, SNAPC4, SNHG7, SNORA17, SNORA43, SSNA1, TMEM141, TMEM203, TRAF2, TUBB2C, UAP1L1, UBAC1, WDR85, ZMYND19,	Loss	
3	10120102	4	Gain	16	396263	1518215	1121.95	160278	**17 genes:** BAIAP3, C16orf42, C16orf91, C1QTNF8, CCDC154, CLCN7, GNPTG, LOC146336, NARFL, NCRNA00235, NHLRC4, PIGQ, PRR25, RHBDL1, SOLH, SSTR5, WDR24	Loss	
6	10092802	3	Gain	1	3701971	4263677	561.706	3955	**4 genes:** C1orf174,DFFB, KIAA0562, LRRC47, LOC100133612	Loss	
7	10100602	3	Gain	1	3726325	3996253	269.928	4907	**4 genes:** DFFB, C1orf174, KIAA0562, LOC100133612,	Loss	
9	10121502	1	Loss	2	242177376	242291694	114.318	977	SEPT2,	Loss	
9	10121501	1	Loss	2	242248524	242273907	25.383	1493	SEPT2,	Loss	
9	10121501	1	Loss	15	25664837	25709733	44.896	1448	UBE3A	Gain	
9	10121501	3	Gain	17	73460617	73702286	241.669	2777	**3 genes:** KIAA0195, CASKIN2, TSEN54	Loss	
10	10120801	3	Gain	1	2088254	4158884	2070.63	8315	**31 genes:** ACTRT2, ARHGEF16, C1orf86, C1orf174, C1orf93, CCDC27, DFFB, FLJ42875, HES5, KIAA0495, KIAA0562, LOC100128003, LOC100129534, LOC100133612, LOC388588, LRRC47, MEGF6, MIR551A, MMEL1, MORN1, PANK4, PEX10, PLCH2, PRDM16, PRKCZ, RER1, SKI, TNFRSF14, TP73, TPRG1L, WDR8,	Loss	601990 Neuroblastoma____(0):NM_005427 (TP73), 602859 Adrenoleukodystrophy_neonatal(202370):NM_002617 (PEX10), Zellweger_syndrome(214100):NM_002617 (PEX10), 164780 1p36_deletion_syndrome___(0):NM_003036 (SKI),
10	10120801	3	Gain	2	241072132	241998576	926.444	23754	**4 genes:** ANKMY1, CAPN10, DUSP28, RNPEPL1	Loss	
10	10120801	3	Gain	9	136087866	140496639	4408.77	56522	**83 genes**: ABCA2, AGPAT2, ANAPC2, C8G, C9orf116, C9orf139, C9orf140, C9orf142, C9orf163, C9orf167, C9orf173, C9orf69, C9orf86, CAMSAP1, CARD9, CLIC3, COBRA1, DNLZ, DPP7, EDF1, EGFL7, ENTPD2, ENTPD8, EXD3, FAM166A, FAM69B, FBXW5, FCN1, FCN2, FUT7, GLT6D1, GPSM1, GRIN1, INPP5E, KIAA0649, KIAA1984, LCN10, LCN12, LCN15, LCN6, LCN8, LCN9, LCNL1, LHX3, LOC100131193, LOC100289341, LOC26102, LRRC26, MAMDC4, MAN1B1, MIR126, MRPL41, MRPS2, NACC2, NDOR1, NELF, NOTCH1, NOXA1, NPDC1, NRARP, OLFM1, PAEP, PHPT1, PMPCA, PNPLA7, QSOX2, RNF208, SDCCAG3, SEC16A, SNAPC4, SNHG7, SNORA17, SNORA43, SSNA1, TMEM141, TMEM203, TRAF2, TUBB2C, UAP1L1, UBAC1, WDR85, ZMYND19,	Loss	
11	10120601	3	Gain	19	4382829	4404831	22.002	523	CHAF1A	Gain	

Cytogenetic array results of selection method 4

Region	File	Type	Chromosome	Min	Max	OMIM entry
1	10100601	LOH	9	93095642	102041615	227645 Fanconi_anemia(0):NM_000136 (FANCC), 601309 Basal_cell_carcinoma_somatic(605462):NM_001083607 (PTCH1), 601309 Basal_cell_nevus_syndrome(109400):NM_000264 (PTCH1), 601309 Holoprosencephaly-7(610828):NM_001083607 (PTCH1), 605573 Pseudohermaphroditism_male_with_gynecomastia(264300):NM_000197 (HSD17B3), 605712 Neuropathy_hereditary_sensory_and_autonomic_type_1(162400):NM_006415 (SPTLC1) , 608135 Lumbar_disc_degeneration(603932):NM_017680 (ASPN), 608135 Osteoarthritis_susceptibility_3(607850):NM_017680 (ASPN), 611570 Fructose-16-bidphosphatase_deficiency(229700):NM_000507 (FBP1),
	10101102	LOH	9	94820892	100203708	
2	10100601	LOH	11	80332710	109251435	602365 Haim-Munk_syndrome(245010):NM_148170 (CTSC), 602365 Papillon-Lefevre_syndrome(245000):NM_148170 (CTSC), 602365 Periodontitis_juvenile(170650):NM_148170 (CTSC), 603025 Leukemia_acute_myeloid(601626):NM_007166 (PICALM), 603025 Leukemia(0):NM_007166 (PICALM), 604579 Retinopathy_of_prematurity(133780):NM_012193 (FZD4), 606933 Albinism_oculocutaneous_type_IA(203100):NM_000372 (TYR), 606933 Skin_hair_eye_pigmentation_3_freckling(601800):NM_000372 (TYR) , 606933
	10101102	LOH	11	80341129	88960147	Waardenburg_syndrome_albinism_digenic(103470):NM_000372 (TYR), 612988 Optic_atrophy-7(612989):NM_032273 (TMEM126A),
3	10100601	LOH	12	113273049	126127471	142410 Diabetes_mellitus_insulin-dependent(222100):NM_000545 (HNF1A), 142410 Diabetes_mellitus_noninsulin-dependent_2(125853):NM_000545 (HNF1A), 142410 Hepatic_adenoma(142330):NM_000545 (HNF1A), 142410 MODY_type_III(600496):NM_000545 (HNF1A), 142410 Renal_cell_carcinoma(144700):NM_000545 (HNF1A), 163731 Pyloric_stenosis_infantile_hypertrophic_1_susceptibility_to(179010):NM_000620 (NOS1), 601406 B-cell_non-Hodgkin_lymphoma(0):NM_020993 (BCL7A), 601620 Holt-Oram_syndrome(142900):NM_181486 (TBX5), 601621 Ulnar-mammary_syndrome(181450):NM_005996 (TBX3), 606686 Leukoencephalopathy_with_vanishing_white_matter(603896):NM_001414 (EIF2B1), 606885 Acyl-CoA_dehydrogenase_short-chain_deficiency_of(201470):NM_000017 (ACADS), 607444 Shwachman-Bodian-Diamond_syndrome(260400):NM_006843 (SDS), 608014 Neuropathy_distal_hereditary_motor_type_IIA(158590):NM_014365 (HSPB8), 609695 Hawkinsinuria(140350):NM_002150 (HPD), 609695 Tyrosinemia_type_III(276710):NM_002150 (HPD), 610277 Immune_dysfunction_with_T-cell_inactivation_due_to_calcium_entry_defect_1(612782):NM_032790 (ORAI1), 611716 Cutis_laxa_autosomal_recessive_type_II(219200):NM_012463 (ATP6V0A2), 611716 Wrinkly_skin_syndrome(278250):NM_012463 (ATP6V0A2),
	10101102	LOH	12	113740466	126481108	
4	10101102	LOH	16	25991284	31947653	103850 Glycogen_storage_disease_XII(611881):NM_184043 (ALDOA), 107265 Antibody_deficiency_due_to_defect_in_CD19___(0):NM_001770 (CD19), 108730 Brody_myopathy(601003):NM_173201 (ATP2A1), 120980 Systemic_lupus_erythematous_association_with_susceptibility_to_6(609939):NM_001145808 (ITGAM), 137070 Amyotrophic_lateral_sclerosis_6_autosomal_recessive(608030):NM_001170937 (FUS), 147781 AIDS_slow_progression_to(609423):NM_000418 (IL4R), 147781 Atopy_susceptibility_to(147050):NM_000418 (IL4R), 157700 Mitral_valve_prolapse(0):NM_017458 (MVP), 172471 Cirrhosis_due_to_liver_phosphorylase_kinase_deficiency___(0):NM_001172432 (PHKG2), 172471 Glycogen_storage_disease_IXc(613027):NM_001172432 (PHKG2), 182381 Renal_glucosuria(233100):NM_003041 (SLC5A2), 602389 Combined_oxidative_phosphorylation_deficiency_4(610678):NM_003321 (TUFM), 605383 IgE_elevated_level_of(147050):NM_181079 (IL21R), 605383 Lymphoma(0):NM_181079 (IL21R), 607042 Ceroid-lipofuscinosis_neuronal-3_juvenile(204200):NM_001042432 (CLN3), 607764 Cholestasis_progressive_familial_intrahepatic_4(607765):NM_025193 (HSD3B7), 608547 Vitamin_K-dependent_clotting_factors_combined_deficiency_of_2(607473):NM_206824 (VKORC1), 608547 Warfarin_resistance(122700):NM_206824 (VKORC1), 137070 Amyotrophic_lateral_sclerosis_6_autosomal_recessive(608030):NM_001170634 (FUS),
	10100601	LOH	16	26280223	31947653	
5	10100601	LOH	16	46540545	78675879	104155 Prostate_cancer_susceptibility_to(176807):NM_001164766 (ZFHX3), 114019 Mental_retardation_autosomal_dominant_3(612580):NM_001793 (CDH3), 114021 Ectodermal_dysplasia_ectrodactyly_and_macular_dystrophy(225280):NM_001793 (CDH3), 114021 Hypotrichosis_congenital_with_juvenile_macular_dystrophy(601553):NM_001793 (CDH3), 121360 Myeloid_leukemia_acute(0):NM_022845 (CBFB), 125860 Benzene_toxicity(0):NM_000903 (NQO1), 125860 Breast_cancer(0):NM_001025434 (NQO1), 125860 Leukemia_post-chemotherapy(0):NM_001025434 (NQO1), 140100 Anhaptoglobinemia___(0):NM_001126102 (HP), 140100 Hypohaptoglobinemia___(0):NM_001126102 (HP), 172490 Phosphorylase_kinase_deficiency_of_liver_and_muscle_autosomal_recessive(261750):NM_001031835 (PHKB), 188250 Mitochondrial_DNA_depletion_syndrome_myopathic_form(609560):NM_004614 (TK2), 192090 Breast_cancer(0):NM_004360 (CDH1), 192090 Cleft_lip_with_or_without_cleft_palate_with_gastric_cancer(0):NM_004360 (CDH1), 192090 Endometrial_carcinoma___(0):NM_004360 (CDH1), 192090 Gastric_cancer_familial_diffuse(137215):NM_004360 (CDH1), 192090 Listeria_monocytogenes(0):NM_004360 (CDH1), 192090 Ovarian_carcinoma___(0):NM_004360 (CDH1), 218030 Apparent_mineralocorticoid_excess(0):NM_000196 (HSD11B2), 218030 Hypertension(0):NM_000196 (HSD11B2), 601065 Charcot-Marie-Tooth_disease_axonal_type_2N(613287):NM_001605 (AARS), 602218 Townes-Brocks_branchiootorenal-like_syndrome(107480):NM_002968 (SALL1), 602218 Townes-Brocks_syndrome(107480):NM_002968 (SALL1), 602311 Leanness(0):NM_001138 (AGRP), 602311 Obesity_late-onset(601665):NM_001138 (AGRP), 602438 Cataract_lamellar(116800):NM_001538 (HSF4), 602438 Cataract_Marner_type(116800):NM_001538 (HSF4), 605131 Esophageal_squamous_cell_carcinoma(133239):NM_130844 (WWOX), 605956 Blau_syndrome(186580):NM_022162 (NOD2), 605956 Crohn_disease_susceptibility_to(266600):NM_022162 (NOD2), 605956 Psoriatic_arthritis_susceptibility_to(607507):NM_022162 (NOD2), 605956 Sarcoidosis_early-onset(609464):NM_022162 (NOD2), 606967 Fish-eye_disease(136120):NM_000229 (LCAT), 606967 Norum_disease(245900):NM_000229 (LCAT), 606979 Congenital_disorder_of_glycosylation_type_IIh(611182):NM_032382 (COG8), 607040 Earwax_wet_dry(117800):NM_032583 (ABCC11), 607456 Cirrhosis_North_American_Indian_childhood_type(604901):NM_032830 (CIRH1A), 610937 Joubert_syndrome_7(611560):NM_001127897 (RPGRIP1L), 610937 Meckel_syndrome_type_5(611561):NM_015272 (RPGRIP1L), 610966 Growth_retardation_developMental_delay_coarse_facies_and_early_death(612938):NM_001080432 (FTO), 611026 Leukodystrophy_dysmyelinating_and_spastic_paraparesis_with_or_without_dystonia(612443):NM_024306 (FA2H), 612051 Spinocerebellar_ataxia_31(117210):NM_001178020 (BEAN), 613018 Tyrosinemia_type_II(277660):NM_000353 (TAT),
	10101102	LOH	16	46540545	79793766	
6	10101102_(0	LOH	X	8243429	12240607	300552 Opitz_G_syndrome_type_I(300000):NM_033290 (MID1), 300808 Ocular_albinism_type_I_Nettleship-Falls_type(300500):NM_000273 (GPR143), 308700 Kallmann_syndrome___(0):NM_000216 (KAL1),
	10100601_(0	LOH	X	8515200	10641757	
7	10100601_(0	LOH	X	35953801	39116739	300481 Chronic_granulomatous_disease_X-linked(306400):NM_000397 (CYBB), 312610 Cone-rod_dystrophy-1(304020):NM_001034853 (RPGR), 312610 Macular_degeneration(0):NM_001034853 (RPGR), 312610 Retinitis_pigmentosa_X-linked_and_sinorespiratory_infections_with_or_without_deafness(300455):NM_001034853 (RPGR), 312610 Retinitis_pigmentosa-3(300029):NM_001034853 (RPGR), 314850 McLeod_syndrome___(0):NM_021083 (XK), 314850 McLeod_syndrome_with_neuroacanthosis___(0):NM_021083 (XK),
	10101102_(0	LOH	X	35955810	39657155	
8	10101102_(0	LOH	X	62018109	67251413	300429 Hyperekplexia_and_epilepsy(300607):NM_015185 (ARHGEF9)
	10100601_(0	LOH	X	62269648	67412413	
9	10101102_(0	LOH	X	138052949	140892452	300746 Thrombophilia_X-linked_due_to_factor_IX_defect(300807):NM_000133 (F9), 313430 Panhypopituitarism_X-linked(312000):NM_005634 (SOX3), 300746 Warfarin_sensitivity(122700):NM_000133 (F9), 313430 Mental_retardation_X-linked_with_isolated_growth_hormone_deficiency(300123):NM_005634 (SOX3), 300746 Hemophilia_B(306900):NM_000133 (F9)
	10100601_(0	LOH	X	138319265	140374229	
10	10101102_(0	LOH	X	146613127	150792888	300120 Hypospadias_2_X-linked(300758):NM_005491 (MAMLD1), 300415 Myotubular_myopathy_X-linked(310400):NM_000252 (MTM1), 300806 Mental_retardation_X-linked_FRAXE_type(309548):NM_002025 (AFF2), 309550 Fragile_X_syndrome(300624):NM_002024 (FMR1), 309550 Fragile_X_tremor_ataxia_syndrome(300623):NM_002024 (FMR1), 309900 Mucopolysaccharidosis_II___(0):NM_006123 (IDS),
	10100601_(0	LOH	X	146628589	150379375	